IRE Transactions

on INFORMATION THEORY

Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing and Utilization.

Volume IT-7 **APRIL, 1961** **Number 2**

Published Quarterly

In This Issue

The Probabilities of Incorrect Dismissal

On the Asymptotic Efficiency of Locally Optimum Detectors

Frequency Difference Between Two Partially Correlated Noise Channels

Complementary Series

Signal Detection by Adaptive Filters

The Coding of Pictorial Data

Optimum (Bayes) Tests for the Detection of Normal Stochastic Signals

PUBLISHED BY THE
Professional Group on Information Theory

IRE Professional Group on Information Theory

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of an annual fee of $4.00.

IRE TRANSACTIONS® on INFORMATION THEORY is published in January, April, July, and October, by the IRE for the Professional Group on Information Theory, at 1 East 79 Street, New York 21, N. Y. In addition to these regular quarterly issues, Special Issues appear from time to time. Responsibility for contents rests upon the authors and not upon the IRE, the Group, or its members. Individual copies of this issue and all available back issues, except Vol. IT-4, may be purchased at the following prices: IRE members (one copy) $2.25, libraries and colleges $3.25, all others $4.50. Annual subscription rate: libraries and colleges, $12.75, non-members $17.00.

IRE Transactions on Information Theory

A Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing and Utilization

Volume IT-7 APRIL, 1961 Number 2
Published Quarterly

TABLE OF CONTENTS

Contributions

PAGE

Effect of Hard Limiting on the Probabilities of Incorrect Dismissal at the Output of an Envelope Detector *P. Bello and W. Higgins* 60

On the Asymptotic Efficiency of Locally Optimum Detectors *Jack Capon* 67

Frequency Differences between Two Partially Correlated Noise Channels *Janis Galejs* 72

Complementary Series *Marcel J. E. Golay* 82

Signal Detection by Adaptive Filters *Edmund M. Glaser* 87

The Coding of Pictorial Data *Joseph S. Wholey* 99

A Note on Singular and Non-Singular Optimum (Bayes) Tests for the Detection of Normal Stochastic Signals in Normal Noise *David Middleton* 105

Correspondence

A Lower Bound for Error Detecting and Error Correcting Codes *Peggy Tang Strait* 114

A Simple Proof of an Inequality of McMillan *Jack Karush* 118

Note on an Integral Equation Occurring in the Prediction, Detection and Analysis of Multiple Time Series *J. B. Thomas and L. A. Zadeh* 118

Contributors 121

Abstracts 122

Book Reviews 125

Effect of Hard Limiting on the Probabilities of Incorrect Dismissal and False Alarm at the Output of an Envelope Detector*

P. BELLO†, ASSOCIATE MEMBER, IRE, AND W. HIGGINS‡, SENIOR MEMBER, IRE

Summary—**This paper is concerned with the effect of hard limiting on the signal detectability of a system consisting of a limiter, narrow-band filter, and envelope detector in cascade. The input to the system is a pulsed IF signal immersed in noise whose power spectrum is uniform over a band of width *W* cycles.**

Assuming that the noise bandwidth *W* is much larger than the bandwidth of the narrow-band filter, the probability distribution of the output of the filter will approach Gaussian. A bivariate Edgeworth series approximation is necessary to handle the narrow-band-filter output since the "in-phase" and "quadrature" components of the narrow-band-filter output are statistically dependent random variables. An expression is derived for the probability of incorrect dismissal that requires the numerical evaluation of single integrals only. From the same bivariate Edgeworth series, an expression is derived for the probability-density function of the output of the envelope detector for the zero-input-signal case. Subsequent integration leads to the probability of false alarm.

Introduction

In the attempt to achieve a constant-false-alarm-rate (CFAR) capability, a limiter is frequently used before signal filtering and detection in a radar system. The effect of the limiter on the probabilities of false alarm and incorrect dismissal is of interest. This paper[1] is concerned with the system of Fig. 1, which shows a limiter followed by a narrow-band filter, envelope detector, and threshold device in cascade. The input to the limiter consists of stationary Gaussian noise plus an IF pulse train. The output of the envelope detector is a train of video pulses perturbed nonlinearly by noise.

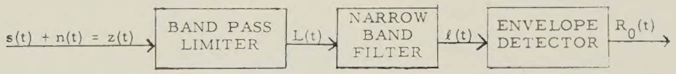


Fig. 1—System to be analyzed.

The following expressions for signal and noise, respectively, apply at the input to the limiter:

$$s(t) = P(t) \cos \omega_0 t$$

$$n(t) = x(t) \cos \omega_0 t - y(t) \sin \omega_0 t. \quad (1)$$

This expression for $s(t)$ presumes a coherent set of pulses. Since the output statistics of the envelope detector are functionally independent of the initial phase of the set of input pulses; and since it will be assumed henceforth that the narrowband filter "integrates" over only one pulse, the assumption of coherence will not matter. The input to the limiter, $z(t)$, is given by

$$\begin{aligned} z(t) &= [x(t) + P(t)] \cos \omega_0 t - y(t) \sin \omega_0 t \\ &= R(t) \cos [\omega_0 t + \phi(t)], \end{aligned} \quad (2)$$

where $R(t)$ is the envelope and $\phi(t)$ is the phase of $z(t)$.

It is assumed that $n(t)$ is a narrowband noise of bandwidth W centered on ω_0 radians per second. The bandpass limiter is assumed to be ideal in the sense that its output $L(t)$ is given by

$$L(t) = \cos [\omega_0 t + \phi(t)]; \quad (3)$$

i.e., envelope variations have been completely removed. $L(t)$ may be expressed in the form

$$L(t) = X(t) \cos \omega_0 t - Y(t) \sin \omega_0 t, \quad (4)$$

where $X(t)$ and $Y(t)$, the amplitudes of the "in-phase" and "quadrature" components, are given by,

$$X(t) = \cos \phi(t) = \frac{x(t) + P(t)}{\sqrt{[x(t) + P(t)]^2 + y^2(t)}},$$

$$Y(t) = \sin \phi(t) = \frac{y(t)}{\sqrt{[x(t) + P(t)]^2 + y^2(t)}}. \quad (5)$$

The output of the narrow-band filter, $l(t)$, may be expressed as

$$l(t) = X_0(t) \cos \omega_0 t - Y_0(t) \sin \omega_0 t, \quad (6)$$

where $X_0(t)$ and $Y_0(t)$ are the amplitudes of the "in-phase" and "quadrature" components of $l(t)$.

If the narrow-band filter is symmetrical about its center frequency ω_0, its impulse response may be expressed in the form

$$h(t) = E(t) \cos \omega_0 t. \quad (7)$$

It may readily be shown that $X_0(t)$ and $Y_0(t)$ are then given (apart from an irrelevant constant multiplier) by

$$\left.\begin{aligned} X_0(t) &= E(t) \otimes X(t) \\ Y_0(t) &= E(t) \otimes Y(t) \end{aligned}\right\}, \quad (8)$$

* Received by the PGIT, October 19, 1959; revised manuscript received November 16, 1960.

† Sylvania Electric Products, Inc., Sylvania Electronic Systems Div., Appl. Res. Lab., Waltham, Mass.

‡ Lincoln Lab., Mass. Inst. Tech., Lexington, Mass.

[1] For a more detailed discussion of the method of approach used in this paper, the reader is referred to P. Bello and W. Higgins, "Effect of Limiting on the Probability of Incorrect Dismissal at the Output of an Envelope Detector," Appl. Res. Memo. 163, Appl. Res. Lab., Sylvania Electric Products, Inc., Waltham, Mass.; March, 1959.

ere the operation $\otimes$ is convolution. The output elope from an ideal envelope detector is then

$$R_0 = \sqrt{X_0^2 + Y_0^2}. \tag{9}$$

The problem has been converted from a narrow-band iation to an equivalent low-frequency one concerning y the amplitudes of the in-phase and quadrature comients X, Y at the output of the limiter, and an equivit low-pass filter whose impulse response is equal to envelope of the impulse response of the narrow-band er, $E(t)$.

For convenience we adjust the pulse train so that a eo pulse has its maximum at $t = 0$ (in the absence of se). Then the quantity that we wish to calculate is the bability that the envelope lies below a specified eshold at $t = 0$.[2]

For $t = 0$, (8) implies

$$\int_0^\infty E(t)X(-t)\,dt = X_0(0) \equiv U, \tag{10}$$

$$\int_0^\infty E(t)Y(-t)\,dt = Y_0(0) \equiv V, \tag{11}$$

ere U and V denote the amplitudes of the "in-phase" l "quadrature" components, respectively, at the row-band-filter output at $t = 0$. The output of the elope detector at $t = 0$ will be denoted by the symbol Thus,

$$D \equiv R_0(0) = \sqrt{U^2 + V^2}. \tag{12}$$

e probability of incorrect dismissal, P_{ID}, is defined as

$$P_{ID} = \Pr[D < B], \tag{13}$$

ere $B > 0$ is some preset threshold level at the output the envelope detector. The probability of false alarm $_A$ is defined as

$$P_{FA} = \Pr[D > B] \quad \text{for} \quad P(t) \equiv 0; \tag{14}$$

, it is the probability that the output of the envelope ector exceeds the threshold in the absence of signal. Because of the nonlinear action of the limiter, its put, $L(t)$ is a non-Gaussian stochastic process. Howr, if the bandwidth of the narrow-band filter is small ugh compared to the bandwidth of $L(t)$ (although we ll not assume it to be so small as to integrate over more n one pulse), then its output $l(t)$ and hence, U and V, be nearly Gaussian. It is assumed that the ratio of the ut-noise bandwidth to the bandwidth of the narrowid filter is much larger than unity. The joint statistics U and V may then be approximated by an Edgeworth es.

n order to make the subsequent mathematics at all table to numerical evaluation, it is necessary to approximate the integrals defining U and V by sums in such a way that only values of the argument separated by $1/W$ are dealt with. When this is done, U and V are each represented as a sum of independent random variables. The conversion of the integrals to sums can be viewed in at least two different ways:

1) A method of numerical integration in which the mesh size is taken as $1/W$.
2) An approximation of the impulse response $E(t)$ of the low-frequency equivalent filter by a series of impulses or its step response by a staircase with jumps occurying every $1/W$ seconds.[3]

Thus U and V will be represented as

$$U = \sum_0^\infty \frac{1}{W} E\left(\frac{k}{W}\right) U_k \qquad V = \sum_0^\infty \frac{1}{W} E\left(\frac{k}{W}\right) V_k, \tag{15}$$

where

$$U_k = X\left(-\frac{k}{W}\right) \qquad V_k = Y\left(-\frac{k}{W}\right). \tag{16}$$

In view of the assumption that the input noise spectrum is uniform, one readily determines that $X(-k/W)$ is independent of $X(-j/W)$; $Y(-k/W)$ is independent of $Y(-j/W)$; and $Y(-j/W)$ is independent of $X(-k/W)$, $j \neq k$. However $Y(-k/W)$ is not independent of $X(-k/W)$.

It will be necessary for later developments to deal with standardized random variables. To this end let us define the standardized variables

$$u = \frac{U - m_U}{\sigma_U}, \qquad v = \frac{V - m_V}{\sigma_V},$$

$$u_k = \frac{U_k - m_{U_k}}{\sigma_{U_k}}, \qquad v_k = \frac{V_k - m_{V_k}}{\sigma_{V_k}}, \tag{17}$$

where the notation m_Q and σ_Q^2 is used to denote the mean and variance of a random variable Q. In terms of standardized variables, (15) becomes

$$u = \sum_0^\infty \alpha_k u_k \qquad v = \sum_0^\infty \beta_k v_k, \tag{18}$$

where

$$\alpha_k = \frac{1}{W}\frac{\sigma_{U_k}}{\sigma_U} E\left(\frac{k}{W}\right) \qquad \beta_k = \frac{1}{W}\frac{\sigma_{V_k}}{\sigma_V} E\left(\frac{k}{W}\right). \tag{19}$$

In the following section the bivariate Edgeworth series expansion of the density function of U and V will be derived.

[2] The maximum of the envelope may not occur at the sampled e instants due to noise fluctuations. However, when the range of a radar is narrow, the model assumed in the analysis is a onable approximation to an actual radar system.

[3] The absolute accuracy to which the integrals representing U and V are approximated by sums is not of prime importance here. What we want to determine in this paper is the *change* in the first order statistics at the envelope detector output caused by the introduction of the limiter. So long as the bandwidth of the narrowband filter is small compared to W, its precise transfer function shape (or impulse response) only weakly affects the output statistics. Thus, the approximation 2 (above) itself might be considered a suitable low frequency equivalent filter for the purpose of this paper.

Edgeworth Series

By a straightforward extension of the univariate Edgeworth series expansion,[4,5] one may determine a bivariate Edgeworth series expansion for the pair of random variables (u, v). To orient the reader, a brief discussion of this extension will be given. Let

$$F(\xi, \eta) = \overline{\exp [i\xi u + i\eta v]}, \; F_k(\xi, \eta) = \overline{\exp [i\xi u_k + i\eta v_k]}, \tag{20}$$

(where the bar denotes a statistical average) denote the characteristic function of the pairs (u, v) and (u_k, v_k) respectively. The Taylor series expansion of the logarithm of $F(\xi, \eta)$ about $\xi = 0$, $\eta = 0$ takes the form

$$\log F(\xi, \eta) = -\frac{\xi^2}{2} - \overline{uv}\xi\eta - \frac{\eta^2}{2} + \sum_{n=0}' \sum_{m=0}' \gamma_{mn} \frac{(i\xi)^m (i\eta)^n}{m!\, n!}, \tag{21}$$

where the primed sums indicate that only terms for $m + n \geq 3$ are to be included in the sum and the semi-invariant γ_{mn} is given by

$$\gamma_{mn} = \frac{\partial^{m+n} lnF(\xi, \eta)}{i^{m+n}\, \partial\xi^m\, \partial\eta^n}. \tag{22}$$

Examination of (21) shows that $F(\xi, \eta)$ may be expressed in the form

$$F(\xi, \eta) = \exp \left[-\frac{\xi^2 + 2\overline{uv}\xi\eta + \eta^2}{2} \right] \cdot \exp \left[\sum' \sum' \gamma_{mn} \frac{(i\xi)^m (i\eta)^n}{m!\, n!} \right]. \tag{23}$$

The first factor in (23) is recognized as the joint characteristic function for a pair of standardized Gaussian random variables. Let the second factor, defined as $G(\xi, \eta)$, be expanded as follows

$$G(\xi, \eta) = \exp \left[\sum' \sum'\right] = \sum_{l=0}^{\infty} \frac{\left[\sum' \sum'\right]^l}{l!}. \tag{24}$$

Use of the expansion of $G(\xi, \eta)$ of (24) in (23) leads to an expansion of $F(\xi, \eta)$ in a series of terms each of which is a product of the Gaussian characteristic function times powers of ξ, η. Inverse Fourier transforming this series term by term leads to an expression for the joint density function of u and v in a series, the leading term of which is the standardized joint normal-probability density function. Subsequent terms involve the derivatives of this latter function weighted by appropriate coefficients involving the semi-invariants γ_{mn}.

This series expansion as such does not become the Edgeworth series expansion until terms of the same "order" are grouped together. A series of terms are defined to be of the same "order" if their coefficients are proportional to the same power of the ratio r of the bandwidth (suitably defined) of the narrow-band filter to the bandwidth of the input noise W. Thus our Edgeworth series is an expansion in powers of the bandwidth ratio.

One may show that the semi-invariant γ_{mn} is of the order $\frac{1}{2}(m + n) - 1$, *i.e.*,

$$\gamma_{mn} \sim r^{(m+n/2)-1}. \tag{25}$$

In the expansion of $G(\xi, \eta)$, only γ_{mn} values for $m + n \geq 3$ are involved. Thus, following the leading (Gaussian) term, the first correction term in the Edgeworth series is proportional to the square root of the bandwidth ratio. The successive higher-order terms are proportional to r, $r^{3/2}$, r^2, etc. To determine all the terms of the same order it is important to note that the "order" of the product of n semi-invariants is equal to the sum of the orders of the individual semi-invariants. If we examine the expansion of $G(\xi, \eta)$ in powers of $\sum'\sum'$ we note that in a term of the form $(\sum'\sum')^n$ only products of semi-invariants taken n at a time are involved. Thus, one may show that all terms of order $\frac{1}{2}$ comes from $\sum'\sum'$, all terms of order 1 come from $\sum'\sum'$ and $(\sum'\sum')^2$, etc. In this way it is possible to collect all terms of the same order.

Without going into the tedious details, we find that the Edgeworth series expansion for terms up to order 1 is

$$W(u, v) = \phi(u, v) - \left\{ \sum_{m=0}^{3} \gamma_{m,3-m} \frac{\phi_{m,3-m}}{m!\,(3 - m)!} \right\} + \left\{ \sum_{m=0}^{4} \gamma_{m,4-m} \frac{\phi_{m,4-m}}{m!\,(4 - m)!} + \frac{1}{2} \sum_{n=0}^{3} \sum_{m=0}^{3} \gamma_{m,3-m}\gamma_{n,3-n} \cdot \frac{\phi_{m+n,6-(m+n)}}{(m!)(3 - m)!\, n!\,(3 - n)!} \right\} + \cdots \tag{26}$$

where

$$\phi(u, v) = \frac{1}{2\pi\sqrt{1 - (\overline{uv})^2}} \exp \left[-\frac{u^2 - 2\overline{uv}uv + v^2}{2[1 - (\overline{uv})^2]} \right], \qquad \phi_{mn}(u, v) = \frac{\partial^{m+n}\phi(u, v)}{\partial u^m\, \partial v^n}, \tag{27}$$

The first term in (26) in braces is of order $\frac{1}{2}$ and the second term is of order 1. One may readily generalize (26) to include terms of arbitrary order, however, the number of terms increases quite rapidly.

It is clear that the desired Edgeworth series may be found once the γ_{mn} are determined. By carrying through the indicated differentiations in (22), one may obtain expressions for γ_{mn} in terms of the moments of the standardized variables u, v. These expressions have been calculated by the authors for $m + n \leq 6$ and are presented in Table I. Only those γ_{mn} for $m \geq n$ are given since γ_{nm} may be obtained from γ_{mn} by interchanging u and v. It should be noted that $\gamma_{00} \equiv 1$ and $\gamma_{01} = \gamma_{10} = 0$,

[4] H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J., p. 227; 1946.

[5] D. Middleton, "Theory of random noise; phenomenological models," *J. Appl. Phys.*, vol. 22, pp. 1153–1163; September, 1951.

TABLE I
SEMI-INVARIANTS IN TERMS OF MOMENTS

$$= \overline{uv}; \quad \gamma_{30} = \overline{u^3}; \quad \gamma_{21} = \overline{u^2v}$$

$$= \overline{u^4} - 3; \quad \gamma_{31} = \overline{u^3v} - 3\overline{uv}; \quad \gamma_{22} = \overline{u^2v^2} - 2(\overline{uv})^2 - 1$$

$$= \overline{u^5} - 10\overline{u^3}; \quad \gamma_{41} = \overline{u^4v} - 4(\overline{u^3})(\overline{uv}) - 6\overline{u^2v}$$

$$= \overline{u^3v^2} - 6(\overline{u^2v})(\overline{uv}) - 3\overline{uv^2} - \overline{u^3}$$

$$= \overline{u^6} - 15\overline{u^4} - 10(\overline{u^3})^2 + 30$$

$$= \overline{u^5v} - 5(\overline{uv})(\overline{u^4}) - 10(\overline{u^2v})(\overline{u^3}) - 10\overline{u^3v} + 30\overline{uv}$$

$$= \overline{u^4v^2} - 8(\overline{uv})(\overline{u^3v}) - 6(\overline{u^2v})^2 - 6\overline{u^2v^2} - 4(\overline{u^3})(\overline{uv^2}) + 24(\overline{uv})^2 - \overline{u^4} + 6$$

$$= \overline{u^3v^3} - 3(\overline{u^3v} + \overline{v^3u}) - 9(\overline{uv^2})(\overline{u^2v}) - 9(\overline{uv})(\overline{u^2v^2}) + 18\overline{uv} - (\overline{u^3})(\overline{v^3}) + 12(\overline{uv})^3$$

$= \gamma_{02} = 1$, due to the standardization of u, v. It should be noted that the relation between the semi-in- iants and moments shown in Table I are valid for any r of standardized random variables. In particular, y apply to the pair (u_k, v_k) in the sums defining u and see (18)]. Let a typical semi-invariant of the pair $, v_k)$ be denoted by γ_{mnk}. Then it is readily deduced m (18) and the assumed independence of the pairs $, v_k)$ and (u_j, v_j), that

$$\gamma_{mn} = \sum_{0}^{\infty} \gamma_{mnk} \alpha_k^m \beta_k^n. \tag{28}$$

becomes clear that evaluation of γ_{mn} depends upon the ermination of the typical moment $\overline{u_k^r v_k^s}$ for $k = 1 \cdots \infty$. [Examination of (19) shows that α_k and β_k y be determined once this typical moment is found.] e evaluation of this moment is discussed in the appendix.

PROBABILITY OF INCORRECT DISMISSAL

The probability of incorrect dismissal will now be ressed in terms of an integration over the density ction of the standardized variables. From (12) and) it is readily deduced that

$$= \Pr[-B \le U \le B; \; -\sqrt{B^2 - U^2} \le V \le \sqrt{B^2 - U^2}]. \tag{29}$$

m (17) it is then found that

$$= \Pr\left[\frac{-B - m_U}{\sigma_U} \le u \le \frac{B - m_U}{\sigma_U}; \; \frac{-\sqrt{B^2 - (u\sigma_U + m_U)^2} - m_V}{\sigma_V} \le v \le \frac{\sqrt{B^2 - (u\sigma_U + m_U)^2} - m_V}{\sigma_V}\right]. \tag{30}$$

From the results shown in the Appendix, one may show that m_V is zero. Consequently, if we define the function $A(u)$ by

$$A(u) = \frac{\sqrt{B^2 - (u\sigma_U + m_U)^2}}{\sigma_V}, \tag{31}$$

it follows that

$$P_{ID} = \int_{-B-m_U/\sigma_U}^{B-m_U/\sigma_U} \int_{-A(u)}^{A(u)} W(u, v)\, du\, dv. \tag{32}$$

Because of our choice of $s(t)$ as indicated in (1) it turns out that

$$\left.\begin{aligned} \gamma_{mn} &= 0 \\ \overline{u^m v^n} &= 0 \end{aligned}\right\} n \quad \text{odd}. \tag{33}$$

Thus, in particular, $\overline{uv} = 0$. This means that instead of having to deal with the general $\phi(u, v)$ in (26) we have the special case $\phi(u, v) = \phi(u)\phi(v)$ where

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right], \tag{34}$$

is the univariate normal probability density function.

If we examine the Edgeworth series approximation to $W(u, v)$ given in (26) we see that application of (32) to determine P_{ID} requires the evaluation of the typical integral

$$I_{mn} = \int_{-B-m_U/\sigma_U}^{B-m_U/\sigma_U} \int_{-A(u)}^{A(u)} \phi_m(u)\phi_n(v)\, du\, dv, \tag{35}$$

where

$$\phi_m(u) = \frac{d^m\phi(u)}{du^m}. \tag{36}$$

The integration with respect to v may be performed, yielding

$$I_{mn} = \begin{cases} 0; & n \text{ odd} \\ 2\displaystyle\int_{-B-m_U/\sigma_U}^{B-m_U/\sigma_U} \phi_m(u)\phi_{n-1}[A(u)]\, du; & n \text{ even}, n \neq 0 \\ \displaystyle\int_{-B-m_U/\sigma_U}^{B-m_U/\sigma_U} \phi_m(u)\{2\phi_{-1}[A(u)] - 1\}\, du; & n = 0 \end{cases} \tag{37}$$

where

$$\phi_{-1}(x) = \int_{-\infty}^{x} \phi(\xi)\, d\xi. \tag{38}$$

The integrals I_{mn} were programmed for numerical evaluation on the ELECOM.

Our expression for P_{ID} including terms up to order 1 becomes

$$P_{ID} = I_{00} - \left\{\frac{\gamma_{30}}{6} I_{30} + \frac{\gamma_{12}}{2} I_{12}\right\} + \left\{\frac{\gamma_{40}}{24} I_{40} + \frac{\gamma_{22}}{4} I_{22} + \frac{\gamma_{04}}{24} I_{04} + \frac{\gamma_{30}^2}{72} I_{60} + \frac{\gamma_{12}}{8} I_{24} + \frac{\gamma_{30}\gamma_{12}}{12} I_{42}\right\} + \cdots. \tag{39}$$

For the numerical results of this paper, the Edgeworth series has been calculated for the specific case in which the narrow-band filter is a single-tuned high-Q circuit, and

$$\frac{\tau}{T} = 0.4\pi; \qquad WT = \frac{150}{\pi}, \tag{40}$$

where τ is the width of a typical input (square) pulse and T is the filter time constant. Curves of P_{ID} are plotted as a function of two parameters, $S/\sigma\sqrt{2}$ and B/σ_N. S is the peak signal input, σ is the rms noise level at the input to the limiter, and σ_N is the rms noise level at the limiter output (in the absence of signal). Thus, $S/\sigma\sqrt{2}$ is the rms signal-to-noise ratio at the limiter input and B/σ_N is the threshold level normalized with respect to the output noise level.

The actual calculations of P_{ID} employed the Edgeworth series approximation for terms up to and including order 1. Thus the Gaussian approximation (the first term) and two correction terms are used. However, it is felt that for the range of values of parameters considered, these three terms give an accurate representation. An idea of the accuracy may be inferred from the fact that the plotted curves of P_{ID} with and without the last correction term are indistinguishable graphically. In Fig. 2, P_{ID} is plotted as a function of $S/\sigma\sqrt{2}$ for different normalized thresholds $B/\sigma_N = 2, 3, 4, 5$. The solid curves apply to the system of Fig. 1, while the dashed curves apply to the same system without the limiter.[6] Discussion of these curves will be deferred until after the following section which takes up the evaluation of P_{FA}.

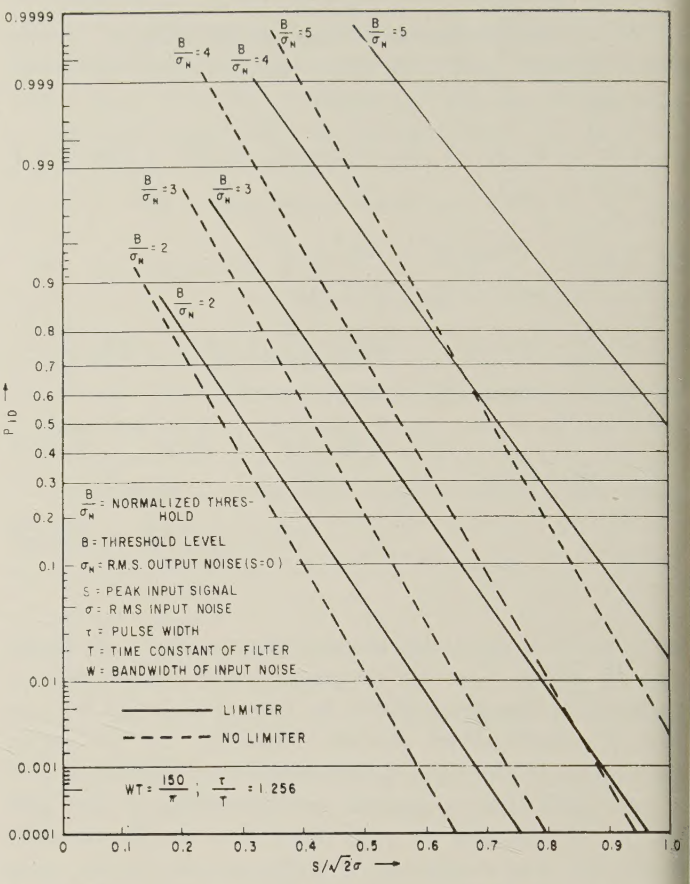


Fig. 2—Probability of incorrect dismissal as a function of signal to noise ratio.

Probability of False Alarm

The probability of false alarm, P_{FA}, has been calculated previously for the system of Fig. 1 on the assumption that U and V are independent random variables.[7] Such an assumption leads to the requirement of only a conventional univariate Edgeworth series. In this section, P_{FA} will be evaluated considering the dependence of U and V. The starting point for this evaluation is the bivariate Edgeworth series expansion of $W(u, v)$ in (26). Considerable simplification results if it is noted that γ_{mn} is zero for m or n odd when $P(t) = 0$, *i.e.*, no input signal. Moreover $\gamma_{mn} = \gamma_{nm}$ for this case. Also when $P(t) = 0$ it is possible to avoid numerical integrations. It will now be demonstrated that both the probability density function of the envelope detector output, and the probability of of false alarm may be expressed in an Edgeworth series involving Laguerre polynomials.

If e and ψ are defined implicitly by

$$\begin{aligned} u &= e \cos \psi \\ v &= e \sin \psi, \end{aligned} \tag{41}$$

then

$$P_{FA} = \Pr\,[D > B] = \Pr\left[e > \sqrt{2}\,\frac{B}{\sigma_N}\right], \tag{42}$$

where B/σ_N is the threshold level in units of rms noise output. Let the joint density function of e and ψ be denoted by $W_1(e, \psi)$ and the density function of e by $W_e(e)$, then

$$W_e(e) = \int_0^{2\pi} W_1(e, \psi)\, d\psi, \tag{43}$$

and

$$P_{FA} = \int_{\sqrt{2}\frac{B}{\sigma_N}}^{\infty} W_e(e)\, de. \tag{44}$$

$W_1(e, \psi)$ is determined from $W(u, v)$ by

$$W_1(e, \psi) = eW(e \cos \psi, e \sin \psi). \tag{45}$$

If it is noted that

$$\phi_n(x) = (-1)^n H_n(x)\phi(x), \tag{46}$$

[6] When the limiter is absent, one may readily determine that

$$P_{ID} = 1 - Q\left[2\left[\frac{S}{\sigma\sqrt{2}}\right](1 - e^{-\tau/T})\sqrt{WT}, \frac{B\sqrt{2}}{\sigma_N}\right],$$

where $Q(\alpha, \beta)$ is the Q function shown in J. I. Marcum, "Tables of Functions RAND Corp., Santa Monica, Calif., Project Rept. RM-339; January 1, 1950.

[7] J. Galejs and J. Storer, "Effects of Limiting on the False Alarm Rate of an Envelope Detector," Appl. Res. Lab., Sylvania Electric Products, Inc., Waltham, Mass., Appl. Res. Memo. 135; June, 1958.

ere $H_n(x)$ is the Hermite polynomial of the nth order, use is made of the integral[8]

$$\int_0^{2\pi} H_{2m}(e \cos \psi) H_{2n}(e \sin \psi)\, d\psi = \frac{(-1)^{m+n}(2n)!\,(2m)!}{2^{m+n} n!\, m!} L_{m+n}\left(\frac{e^2}{2}\right), \qquad (47)$$

ere $L_n(x)$ is the nth order Laguerre polynomial $(x) = 1 - 2x + x^2/2)$, then one finds (after going ough the tedium of evaluating the semi-invariants) t

$$(e) = e \exp\left[-\frac{e^2}{2}\right]\left\{1 - \left(\frac{1}{2WT}\right)L_2\left(\frac{e^2}{2}\right) - \left(\frac{1}{WT}\right)^2\left[\frac{8}{9} L_3\left(\frac{e^2}{2}\right) - \frac{9}{16} L_4\left(\frac{e^2}{2}\right)\right] \cdots\right\} \qquad (48)$$

he probability of false alarm may be calculated with aid of the integral

$$\sqrt{2}\frac{B}{\sigma_N} e L_n\left(\frac{e^2}{2}\right) \exp -\left[\frac{e^2}{2}\right] de = e^{-(B/\sigma_N)^2}\left[L_n\left(\frac{B^2}{\sigma_N^2}\right) - L_{n-1}\left(\frac{B^2}{\sigma_N^2}\right)\right]. \qquad (49)$$

he probabilities of false alarm with and without the iter are plotted in Fig. 3 using terms up to the second er. This was done in order to have at least two correc- terms in the Edgeworth series (the terms of order are identically zero in the zero signal case).

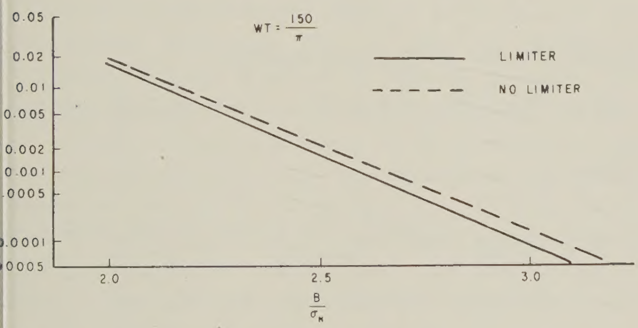


Fig. 3—Probability of false alarm vs normalized threshold.

Discussion

rom the curves of Fig. 2, it may be observed, as might expected, that the introduction of the limiter causes apparent degradation in performance: the probability ncorrect dismissal is higher with the limiter than out it *for the same value of normalized threshold.* A sure of this apparent degradation is the db increase ignal-to-noise radio needed to achieve the same P_{ID} without the limiter (all other parameters being fixed). ection of the curves of Fig. 2, for example, shows that the increase required is approximately 2.7, 2.1, 1.6, and 1.3 db for $(B/\sigma_N) = 5, 4, 3$, and 2, respectively. Such values of degradation may be misleading, however, since it cannot be said that the limiter has hurt signal detectability performance unless it can be demonstrated that the limiter causes a larger P_{ID} for the *same input* signal strength and *the same probability of false alarm* (P_{FA}). It should be noted that while the curves of Fig. 2 allow a comparison of the P_{ID} with and without the limiter on the basis of the same value of input signal strength, they *do not* allow a comparison in addition on the basis of the same P_{FA}. Rather they represent a comparison on the basis of the same normalized threshold. Consequently, one should not jump to the conclusion that the limiter has harmed performance by the degree suggested in Fig. 2 until the P_{FA} with the limiter has been examined as a function of the normalized threshold.

P_{FA} is plotted in Fig. 3 as a function of the normalized threshold for the cases with and without the limiter. In this figure, the bandwidth ratio is 150, and terms of order up to and including $(1/WT)^2$ have been used in the Edgeworth series. These curves were plotted for values of (B/σ_N) in the range $2 \leq (B/\sigma_N) \leq 3$. For values of $(B/\sigma_N) < 2$, the P_{FA} with and without the limiter are essentially identical. For values of (B/σ_N) to any extent greater than 3, P_{FA} drops so rapidly that the first three terms of the Edgeworth series are no longer sufficiently accurate to represent P_{FA} for the case when the limiter is present.

It is clear from the curves that P_{FA} is less when the limiter is used (for the same normalized threshold). However, it is also clear that a very small percentage reduction in threshold is required when the limiter is introduced to maintain the same probability of false alarm as without the limiter, at least for $0 \leq (B/\sigma_N) \leq 3$, and perhaps for values somewhat in excess of 3.

For $(B/\sigma_N) = 3$, $P_{FA} \sim 10^{-4}$ with the limiter. Thus for probabilities of false alarm $\leq 10^{-4}$, it may be said that the introduction of the limiter necessitates an increase in signal strength of approximately 1.6 db, at most $(WT = 150/\pi, (S/2\sigma)^2 < 1)$ to maintain the same probability of incorrect dismissal as without the limiter (and with the same P_{FA}). For $(B/\sigma_N) = 5$, the P_{FA} without the limiter is approximately 10^{-11}, and the P_{FA} with the limiter is even less. Exactly how much less is not known at this point, and to reach an answer would require considerable additional work. However, it is clear that as long as the P_{FA} with the limiter is less than the P_{FA} without the limiter for the same normalized threshold, a comparison of P_{ID} with and without the limiter for the same value of normalized threshold will always provide an upper bound to the degrading effect of the limiter. Thus, we may say that for probabilities of false alarm between 10^{-11} and 10^{-4}, the introduction of the limiter necessitates an increase in signal strength of at most 3 db $(WT = 150/\pi, (S^2/2\sigma^2) < 1)$ to maintain the same probability of incorrect dismissal as without the limiter and with the same P_{FA}.

[8] *Ibid.*, p. A-10, Eq. (30).

Appendix
Calculation of Moments

The typical moment $\overline{u_k^r v_k^s}$ may be expressed in terms of moments of the unstandardized variables as follows

$$\overline{u_k^r v_k^s} = \overline{\left[\frac{U_k - m_{U_k}}{\sigma_{U_k}}\right]^r \left[\frac{V_k - m_{V_k}}{\sigma_{V_k}}\right]^s}$$
$$= \frac{1}{[\sigma_{U_k}]^r [\sigma_{V_k}]^s} \sum_{q=0}^{s} \sum_{p=0}^{r} \begin{bmatrix} r \\ p \end{bmatrix} \begin{bmatrix} s \\ q \end{bmatrix} \overline{U_k^p V_k^q} m_{U_k}^{r-p} m_{V_k}^{s-p}, \qquad (50)$$

where $\begin{bmatrix} m \\ n \end{bmatrix}$ is the combination of m things taken n at a time. From (16) and (5),

$$\overline{U_k^p V_k^q} = \{\cos^p \phi \sin^q \phi\}_{t=(-k/W)}. \qquad (51)$$

It is not difficult to see (and it will be shown subsequently) that the moment in braces in (51) is a function of the ratio $P(t)/\sigma\sqrt{2}$, where $P(t)$ is the envelope of the input pulse train. Thus, we define

$$M_{pq}\left[\frac{P(t)}{\sigma\sqrt{2}}\right] = \overline{\cos^p \phi \sin^q \phi}. \qquad (52)$$

The moment M_{pq} in (52) may be computed by averaging with respect to the joint density function of R and ϕ[9] [see Equation (2)]

$$M_{pq} = \int_0^{2\pi} \int_0^{\infty} \frac{R \cos^p \phi \sin^q \phi}{2\pi\sigma^2} \cdot \exp\left[-\frac{R^2 + P^2 - 2PR\cos\phi}{2\sigma^2}\right] dR\, d\phi. \qquad (53)$$

Since the joint density function of R and ϕ is even in ϕ while $\cos^r \phi \sin^s \phi$ is odd in ϕ for s odd, it is clear that $M_{pq} = 0$ for q odd. For q even, $\cos^p \phi \sin^q \phi$ can be expanded in a finite Fourier cosine series of the form

$$\cos^p \phi \sin^q \phi = \sum_{j=0}^{l} A_j \cos(p + q - 2j)\phi, \qquad (54)$$

where $l = (p + q)/2$ for $p + q$ even, and $l = (p + q - 1)/2$ for $p + q$ odd. With the aid of (53) and the two integrals[10]

$$\frac{1}{2\pi}\int_0^{2\pi} \cos q\phi \exp\left[\frac{PR}{\sigma^2}\cos\phi\right] d\phi = I_q\left(\frac{PR}{\sigma^2}\right),$$

$$\int_0^{\infty} \frac{R}{\sigma^2} I_q\left(\frac{PR}{\sigma^2}\right) \exp\left[-\frac{R^2 + P^2}{2\sigma^2}\right] dR$$
$$= \left[\frac{P}{\sqrt{2}\sigma}\right]^q \frac{\Gamma\left(\frac{q}{2} + 1\right)}{\Gamma(q + 1)} {}_1F_1\left[\frac{q}{2}, q + 1, -\frac{P^2}{2\sigma^2}\right], \qquad (55)$$

it is readily shown that

$$M_{pq}\left[\frac{P}{\sqrt{2}\sigma}\right] = \sum_{j=0}^{l} A_j \left[\frac{P}{\sqrt{2}\sigma}\right]^{p+q-2j} \frac{\Gamma\left[\frac{p+q}{2} - j + 1\right]}{\Gamma(p + q - 2j + 1)}$$
$$\cdot {}_1F_1\left[\frac{p+q}{2} - j, p + q - j + 1, -\frac{P^2}{2\sigma^2}\right], \qquad (56)$$

where ${}_1F_1(\alpha, \beta, x)$ is the confluent hypergeometric function[11] and $\Gamma(x)$ is the gamma function.

According to assumptions previously made, the input pulse train envelope, $P(t)$, is given by

$$P(t) = \begin{cases} S & \text{for } mT_0 - \tau < t < mT_0 \\ 0 & \text{for } (m-1)T_0 < t < mT_0 - \tau, \end{cases} \qquad (57)$$

where T_0 is the period of the pulse train and m is an arbitrary integer. Since we are examining the envelope detector output at $t = 0$, and since we have assumed that T_0 is large enough so that the narrow-band filter integrates over only one pulse, we may take

$$P\left(\frac{-k}{W}\right) = \begin{cases} S & \text{for } 0 < k < K \\ 0 & \text{for } k > K, \end{cases} \qquad (58)$$

where K is the integer that satisfies

$$K = \underset{r}{\text{Max}}\,\{W\tau - r > 0\}; \qquad r = 0, 1, 2, \cdots \qquad (59)$$

Thus,

$$\overline{U_k^p V_k^q} = \begin{cases} M_{pq}\left(\frac{S}{\sigma\sqrt{2}}\right) & \text{for } 0 < k < K \\ M_{pq}(0) & \text{for } k > K. \end{cases} \qquad (60)$$

[9] W. B. Davenport, Jr., and W. L. Root, "Random Signals and Noise," McGraw-Hill Book Co., Inc., New York, N. Y., p. 166, Eq. (8-114); 1958.

[10] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 24, p. 173; 1950.

[11] D. Middleton and V. Johnson, "A Tabulation of Selected Confluent Hypergeometric Functions," Cruft Lab., Harvard University, Cambridge, Mass., Tech. Rept. No. 140; January 5, 1952.

On the Asymptotic Efficiency of Locally Optimum Detectors*

JACK CAPON†, MEMBER, IRE

ummary—A detector examines an unknown waveform to rmine whether it is a mixture of signal and noise, or noise alone. Neyman-Pearson detector is optimum in the sense that for n false alarm probability, signal-to-noise ratio, and number bservations, it minimizes the false dismissal probability. This ctor is optimum for all values of the signal-to-noise ratio, its implementation is usually quite complicated.

n many situations it is desired to detect signals which are very k compared to the noise. The locally optimum detector is ned as one which has optimum properties only for small signal-oise ratios. It is proposed as an alternative to the Neyman-rson detector, since in practice it is usually only necessary to e a near-optimum detector for weak signals, since strong signals be detected with reasonable accuracy even if the detector is below optimum.

n order to evaluate the performance of the locally optimum ctor, it is compared to the Neyman-Pearson detector. This parison is based on the concept of asymptotic relative efficiency oduced by Pitman for comparing hypothesis testing procedures. the basis of this comparison, it is shown that the locally optimum ctor is asymptotically as efficient as the Neyman-Pearson ctor.

number of applications to several detection problems are con-red. It is found that the implementation of the locally optimum ctor is less, or at most as complicated as that of the Neyman-rson detector.

Introduction

THE FUNCTION of a detector is to examine an unknown waveform $Z(t)$ in order to determine whether or not a signal is present in noise. We ıme that the detector's decision is based on the samples $\cdots, Z_n, Z_i = Z(t_i), i = 1, \cdots, n$. If we consider that) is a sample function from the continuous parameter hastic process $\{Z(t)\}$, then $Z_1, \cdots, Z_n$ is a set of dom variables. If we assume that these random iables are mutually independent, and that $\{Z(t)\}$ is ionary, then $Z_1, \cdots, Z_n$ is a set of independent and tically distributed random variables.

hus, the particular detection problem that we are sidering is equivalent to a determination of whether cumulative distribution function (*cdf*) of Z_i, $i = 1$, , n, is $G_\theta(z)$ (signal is present), or is $G_0(z)$ (signal is ent). The parameter θ is the signal-to-noise ratio, and es the purpose of indexing, or labeling, the *cdf* $G_\theta(z)$, we have one *cdf* $G_\theta(z)$ for each θ. In general, θ may be itive or negative; *e.g.*, if we are interested in the problem of detecting a constant signal in additive noise, θ is equal to the ratio of the amplitude of the signal to the rms value of the noise, and will be positive or negative, depending, respectively, on whether the amplitude of the constant signal is positive or negative.

The errors committed by the detector are of the following two exhaustive and mutually exclusive types: a) the detector decides that a signal is present when in reality the signal is absent; the probability of such an error is denoted by α_n, and is known as the false alarm probability; b) the detector decides that a signal is absent when in reality the signal is present; the probability of this type of error is denoted by $\beta_n(\theta)$, and is known as the false dismissal probability. The dependence of this probability on θ is shown explicitly for the purposes of our subsequent discussions.

The Neyman-Pearson detector [1], [2], [3], [13], is optimum in the sense that for given α_n and n, it minimizes $\beta_n(\theta)$, for all θ. Any other fixed-sample detection method must have a larger value of $\beta_n(\theta)$, for each θ, and for the same α_n, n, as the Neyman-Pearson detector. We observe that the Neyman-Pearson detector is optimum for all values of the signal-to-noise ratio. In general, the structure of the Neyman-Pearson detector depends on the input signal-to-noise ratio θ, and changes its form as θ changes. As a consequence, the implementation of this detector, in certain detection problems, is quite complicated, as we shall see subsequently.

In many situations it is desired to detect signals which are very weak compared to the noise. Hence θ will be very close to zero. It would be desirable in these situations to design a detector which has optimum properties only for small signal-to-noise ratios. This detector could also be used for larger signals. This can often be justified in practice by the idea that it is only necessary to have a near-optimum detector for weak signals, since strong signals will be detected even if the detector is well below optimum.

In the present work we give a criterion for determining a detector which has good properties for detecting weak signals in noise. This detector is known as the locally optimum detector. Under certain weak regularity conditions for $G_\theta(z)$, the locally optimum detector is usually much simpler to implement than the Neyman-Pearson detector, and in a certain sense is just as efficient. We shall see subsequently that the concept of the locally optimum detector is related to some previous work [13] on the threshold detection of signals in noise, although the methods used here are quite different from those used previously.

Received by the PGIT, March 28, 1960. This work is based esults which were obtained in a dissertation submitted in partial llment of the requirements for the Ph.D. degree in electrical neering at Columbia University, New York, N. Y. This dis-ation was supported by the USAF under Contract No. AF 04)-4140, and monitored by the Office of Scientific Research, Res. and Dev. Command.
Federal Sci. Corp., New York, N. Y.

The Locally Optimum Detector

If $\beta_n(\theta)$ denotes the false dismissal probability of a detector whose false alarm probability is α_n, then we say that the detector with false alarm probability $\beta_n^*(\theta)$ is the locally optimum detector, for $\theta > 0$, if

$$\frac{\partial}{\partial\theta}\beta_n^*(\theta)\bigg|_{\theta=0} \leq \frac{\partial}{\partial\theta}\beta_n(\theta)\bigg|_{\theta=0}, \quad \text{uniformly in } n,$$

and is the locally optimum detector, for $\theta < 0$, if

$$\frac{\partial}{\partial\theta}\beta_n^*(\theta)\bigg|_{\theta=0} \geq \frac{\partial}{\partial\theta}\beta_n(\theta)\bigg|_{\theta=0}, \quad \text{uniformly in } n.$$

Thus the locally optimum detector minimizes, or maximizes, the slope at $\theta = 0$ of $\beta_n(\theta)$, depending, respectively, on whether θ is greater or less than zero. This definition for the locally optimum detector is similar to a definition given by Lehmann [4] for locally most powerful rank tests.

Before we proceed to determine the structure, or form, of the locally optimum detector we shall state certain regularity conditions which will be required in our subsequent work.

Regularity Conditions

(*i*) The *cdf* $G_\theta(z)$, the probability density function (*pdf*) $g_\theta(z)$ $(= \partial\, G_\theta(z)/\partial z)$, and $\partial\, g_\theta(z)/\partial\theta$ are continuous in the region $-\infty < z < \infty$, $-a \leq \theta \leq a$, $a > 0$, for almost all z; there exist functions $M_0(z)$ and $M_1(z)$, integrable over $(-\infty, \infty)$, such that

$$g_\theta(z) \leq M_0(z), \qquad \left|\frac{\partial g_\theta(z)}{\partial\theta}\right| \leq M_1(z), \qquad -a \leq \theta \leq a.$$

The false alarm probability of a detector is given by

$$\alpha_n = \int \underset{I}{\cdots} \int g_0(z_1) \cdots g_0(z_n)\, dz_1 \cdots dz_n, \tag{1}$$

where I is that region of the n-dimensional sample space of $Z_1, \cdots, Z_n$ for which the detector decides that there is a signal present. Hereafter we refer to the region I as the critical region.

The false dismissal probability $\beta_n(\theta)$ is

$$\beta_n(\theta) = \int \underset{I'}{\cdots} \int g_\theta(z_1) \cdots g_\theta(z_n)\, dz_1 \cdots dz_n, \tag{2}$$

where I' is that region of the n-dimensional sample space of $Z_1, \cdots, Z_n$ which is not included in the critical region. We have

$$\frac{\partial}{\partial\theta}\beta_n(\theta) = \frac{\partial}{\partial\theta}\int \underset{I'}{\cdots} \int g_\theta(z_1) \cdots g_\theta(z_n)\, dz_1 \cdots dz_n. \tag{3}$$

As a consequence of the regularity condition (*i*) we can interchange differentiation and integration [5] in (3) to obtain

$$\frac{\partial}{\partial\theta}\beta_n(\theta)\bigg|_{\theta=0} = \int \underset{I'}{\cdots} \int \frac{\partial}{\partial\theta}[g_\theta(z_1) \cdots g_\theta(z_n)]\, dz_1 \cdots dz_n\bigg|_{\theta=0}. \tag{4}$$

Hence, the problem of determining the locally optimum detector, when $\theta < 0$, is that of choosing the critical region so that the integral in (4) is maximized, subject to the constraint in (1); when $\theta > 0$, the problem is equivalent to choosing the critical region so as to minimize the integral in (4), subject to the constraint in (1). As pointed out by Lehmann [4] we may solve both of these variational problems by means of a direct application of the Neyman-Pearson fundamental lemma [6]. Thus, when $\theta > 0$, the critical region is that region of the n-dimensional sample space of $Z_1, \cdots, Z_n$, for which

$$\frac{\dfrac{\partial}{\partial\theta}\displaystyle\prod_{i=1}^{n} g_\theta(z_i)\bigg|_{\theta=0}}{\displaystyle\prod_{i=1}^{n} g_0(z_i)} > c$$

or

$$L_n(z_1, \cdots, z_n) = \frac{1}{n}\sum_{i=1}^{n} b(z_i) > c \tag{5}$$

where

$$b(z) = \frac{\partial}{\partial\theta}\ln g_\theta(z)\bigg|_{\theta=0}. \tag{6}$$

In an analogous manner we obtain that when $\theta < 0$ the critical region is given by

$$L_n(z_1, \cdots, z_n) < c. \tag{7}$$

The constant c is chosen so as to make the false alarm probability equal to α_n, and is not necessarily the same from one line to the next.

We see that the locally optimum detector is quite simple in structure. It consists of a device which sums a certain function, $b(z)$, of the observations, and a threshold comparator. If this sum exceeds a certain threshold, when $\theta > 0$, the detector decides that the signal is present; otherwise, the decision is that the signal is absent. If this sum is less than a certain threshold, when $\theta < 0$, the detector decides that the signal is present; otherwise, the decision is that the signal is absent.

It should be pointed out that the results obtained above are quite similar to those obtained previously [13] for the threshold-signal design of a detector. In this approach, the design of the detector is based on the leading term of a series expansion for the likelihood ratio taken in powers of the signal-to-noise ratio around zero signal. It is easily seen that this previous approach and the present one yield the same detector. However, in the previous methods it usually is not clear what optimum properties are possessed by the locally optimum detector. In the present

k this detector is shown to be optimum in the sense t it minimizes or maximizes the slope at $\theta = 0$ of θ), depending, respectively, on whether θ is greater or than zero. In addition, in the previous methods it is always clear under what conditions the higher-order ms in θ may be neglected. In the present approach se terms drop out in a natural way. Another important perty of the locally optimum detector, which will be cussed subsequently, is that in a certain sense it is mptotically as efficient as the strictly optimum yman-Pearson detector.

Asymptotic Relative Efficiency of Detection Procedures

n order to evaluate the performance of the locally imum detector we shall compare it to the best possible ector for fixed α_n, namely the Neyman-Pearson detor. This comparison is based on the concept of asymptic relative efficiency (ARE) due to Pitman [7], [8].

Let us suppose that we have two detectors which are igned to detect the same signal, with the same error babilities; suppose further, that the detectors require nple sizes n_1 and n_2, respectively, in order to detect signal with the required error probabilities. Then, $n_1 < n_2$ we would be justified in saying that the first ector is more "efficient" than the second, and would ose the first detector over the second. The criterion oose the detector which requires the smaller sample for the same θ and error probabilities," is, roughly aking, the basis for the concept of ARE. The fact that s concept is useful in comparing detection procedures been pointed out previously [9], [10]. We shall now ke our previous remarks more precise, and give a orous definition for ARE.

Let $\{\theta_i\}$ be a sequence of signal-to-noise ratios such t $\operatorname{limit}_{i\to\infty} \theta_i = 0$, and consider two sequences of detec- procedures $\{D_n\}$, $\{D_n^*\}$, with false dismissal prob- ities $\beta_n(\theta_i)$, $\beta_n^*(\theta_i)$, and the same false alarm prob- ity α. Also let $\{n_i\}$, $\{n_i^*\}$ be two increasing sequences ntegers such that

$$\operatorname*{limit}_{i\to\infty} \beta_{n_i}(\theta_i) = \operatorname*{limit}_{i\to\infty} \beta^*_{n_i^*}(\theta_i), \tag{8}$$

h the two limits existing and not equal to either zero one. Then the ARE of $\{D_n\}$ with respect to $\{D_n^*\}$ is ned as

$$\text{ARE}\ (\{D_n\}, \{D_{n^*}^*\}) = \operatorname*{limit}_{i\to\infty} \frac{n_i^*}{n_i}, \tag{9}$$

his limit exists the same for all sequences $\{n_i\}$, $\{n_i^*\}$ sfying (8). If we denote the statistics on which the ectors $\{D_n\}$, $\{D_n^*\}$ are based as $\{W_n\}$, $\{W_n^*\}$, then the E is denoted as E_{W,W^*} and is given by

$$E_{W,W^*} = \operatorname*{limit}_{i\to\infty} \frac{n_i^*}{n_i}. \tag{10}$$

We will now point out that subject to some regularity conditions, there is a simple expression for the ARE of sequences of detection procedures. We assume that the following conditions are true in some neighborhood of $\theta = 0$: a) $(W_n - E_\theta(W_n))/\sigma_\theta(W_n)$ is asymptotically normal with mean zero and variance one, where $E_\theta(W_n)$ and $\sigma_\theta(W_n)$ are, respectively, the expected value and the standard deviation of W_n, taken under the hypothesis that the *cdf* of Z_i, $i = 1, \cdots, n$, is $G_\theta(z)$; b) for the sequence $\{\theta_n\}$, where $\theta_n = kn^{-1/2}$, k is a constant, we have

$$\operatorname*{limit}_{n\to\infty} \frac{\sigma_{\theta_n}(W_n)}{\sigma_0(W_n)} = 1$$

and

$$E_W = \operatorname*{limit}_{n\to\infty} \left\{\frac{E_{\theta_n}(W_n) - E_0(W_n)}{\theta_n n^{1/2}\sigma_0(W_n)}\right\}^2 = \operatorname*{limit}_{n\to\infty} \left\{\frac{\frac{\partial}{\partial\theta} E_\theta(W_n)}{n^{1/2}\sigma_0(W_n)}\right\}^2_{\theta=0}$$

exists, and is independent of k.

The quantity E_W has been termed the efficacy of the detection procedure based on the sequence of statistics $\{W_n\}$. If conditions a) and b) are satisfied for the detectors $\{D_n\}$, $\{D_{n^*}^*\}$, then it can be shown [8] that

$$E_{W,W^*} = \frac{E_W}{E_{W^*}}. \tag{11}$$

Thus, the ARE of sequences of detection procedures is given by the ratio of the efficacies of the detection procedures.

Asymptotic Efficiency of the Locally Optimum Detector

It follows from the optimum character of the Neyman-Pearson detector that the ARE of any sequence of detectors with respect to the sequence of Neyman-Pearson detectors must be less than or equal to unity. In particular, the efficacy of the Neyman-Pearson detector is greater than or equal to that of any other detection procedure. We now compute this efficacy.

It is well known that the Neyman-Pearson detector bases its decision on the quantity known as the likelihood ratio

$$L_n^*(z_1, \cdots, z_n, \theta) = \prod_{i=1}^{n} \frac{g_\theta(z_i)}{g_0(z_i)}. \tag{12}$$

Since the logarithm is a strictly increasing function of its argument, the Neyman-Pearson detector can base its decision on the quantity $\ln L_n^*$, given by

$$\ln L_n^*(z_1, \cdots, z_n, \theta) = \sum_{i=1}^{n} \ln \frac{g_\theta(z_i)}{g_0(z_i)}. \tag{13}$$

If θ is sufficiently small, and the regularity condition (i) is satisfied, the Neyman-Pearson detector can base its decision on

$$L_n'(z_1, \cdots, z_n, \theta) = \frac{1}{n} \sum_{i=1}^{n} (b(z_i) + o(1)) \tag{14}$$

where

$$\operatorname*{limit}_{\theta \to 0} o(1) = 0, \quad \text{uniformly in} \quad z_1, \cdot \cdot, z_n. \tag{15}$$

Thus, when $\theta > 0$ the Neyman-Pearson detector decides that the signal is present when L_n' exceeds a certain threshold value; otherwise, the decision is that the signal is absent. When $\theta < 0$ the Neyman-Pearson detector decides that the signal is present when L_n' is less than a certain threshold value; otherwise, the decision is that the signal is absent. The threshold value is, of course, chosen in each case so that the false alarm probability of the detector is equal to the prescribed value. We note the similarity of the test statistic L_n' to L_n [*cf.* (14) and (5)].

We see from (14) that, except for the $o(1)$ term, $L_n'(Z_1, \cdots, Z_n, \theta_n)$ is equal to a sum of independent and identically distributed random variable. If we assume that the variance of the random variable $b(Z)$ is bounded, then we have from the central limit theorem that $L_n'(Z_1, \cdots, Z_n, \theta_n)$ is asymptotically normal, so that condition (a) is satisfied. We also obtain from (14) that

$$E_\theta(L_n'(Z_1, \cdots, Z_n, \theta_n)) = E_\theta\left[\frac{\partial}{\partial\theta} \ln g_\theta(Z)\Big|_{\theta=0}\right] + o_n(1), \tag{16}$$

where

$$\operatorname*{limit}_{n \to \infty} o_n(1) = 0. \tag{17}$$

If we differentiate (16) with respect to θ we obtain

$$\frac{\partial}{\partial\theta} E_\theta(L_n'(Z_1, \cdots Z_n, \theta_n))\Big|_{\theta=0} = \inf_{G_0} + o_n(1), \tag{18}$$

where

$$\inf_{G_0} = E_0(b^2(Z)). \tag{19}$$

The quantity defined in (19) is known as the information [14] of the *cdf* $G_\theta(z)$ evaluated at $\theta = 0$, and in our discussions is assumed to be finite. We have from (14) that

$$\sigma_0^2(L_n'(Z_1, \cdots, Z_n, \theta_n)) = \frac{1}{n} (\inf_{G_0} + o_n(1)). \tag{20}$$

Hence, the efficacy of the detection procedure based on the sequence of statistics $\{L_n'\}$ is obtained from (18) and, (20) as

$$E_{L'} = \inf_{G_0}. \tag{21}$$

We shall say that a detection procedure is asymptotically efficient if its efficacy achieves the upper bound in (21), namely $\inf_{G_0}$. We will now show that the locally optimum detector is asymptotically efficient.

We see from (5) that L_n is equal to a sum of independent and identically distributed random variables, each with finite variance. Hence, we have from the central limit theorem that L_n is asymptotically normal, so that condition a) is again satisfied. We also have from (5) that

$$\frac{\partial}{\partial\theta} E_\theta(L_n(Z_1, \cdots, Z_n))\Big|_{\theta=0} = \inf_{G_0} \tag{22}$$

and

$$\sigma_0^2(L_n(Z_1, \cdots, Z_n)) = \inf_{G_0} \tag{23}$$

so that the efficacy of the detection procedure based on the sequence of statistics $\{L_n\}$ is

$$E_L = \inf_{G_0}. \tag{24}$$

Hence the locally optimum detector is asymptotically efficient; *i.e.*, $E_{L,L'} = 1$.

Applications

We now consider a number of detection problems in which our results may be applied. Straightforward calculations show that in each case the regularity conditions are satisfied. We observe that in each example the structure of the locally optimum detector is less, or at most as complicated, as that of the Neyman-Pearson detector.

Detection of a Constant Signal in Additive Gaussian Noise

In this case the signal has a constant amplitude of A, and the noise is normally distributed with mean u and variance s^2. Thus, the *pdf* $g_\theta(z)$ is

$$g_\theta(z) = (2\pi s^2)^{-1/2} \exp\left(-\frac{1}{2}\left(\frac{z-u}{s} - \theta\right)^2\right) \tag{25}$$

where $\theta = A/s$ is the peak signal-to-rms noise ratio. The function $b(z)$ is

$$b(z) = \frac{z-u}{s}, \tag{26}$$

so that the locally optimum detector is based on the statistic

$$\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Z_i - u}{s}\right). \tag{27}$$

If we form the likelihood ratio and take its logarithm, we find that the Neyman-Pearson detector is also based on the statistic in (27), for all $\theta \neq 0$. Thus, in this case the locally optimum detector and the Neyman-Pearson detector coincide. We note that the structure of the Neyman-Pearson detector is the same for all values of $\theta \neq 0$. This is one of those rare examples in which a uniformly optimum detector exists.

ection of a Gaussian Signal in Independent Additive Gaussian Noise

n this example the signal is normally distributed with an zero and variance r^2, and the noise is normally tributed with mean zero and variance s^2. The *pdf* $g_\theta(z)$

$$g_\theta(z) = (2\pi s^2(1 + \theta))^{-1/2} \exp(-\tfrac{1}{2}z^2/(s^2(1 + \theta))) \quad (28)$$

ere $\theta = r^2/s^2$ is the mean-square signal-to-mean-square ise ratio. The function $b(z)$ is

$$b(z) = \frac{1}{2}\left(\frac{z^2}{s^2} - 1\right), \quad (29)$$

that the locally optimum detector is based on the tistic

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Z_i^2}{s^2} - 1\right). \quad (30)$$

If we form the likelihood ratio and take its logarithm, find that the Neyman-Pearson detector is also based the statistic in (30) for all $\theta > 0$, so that it coincides th the locally optimum detector. Since the structure of e Neyman-Pearson detector is the same for all $\theta > 0$, also have a uniformly optimum detector in this case.

velope Detection of a Sine Wave in Narrow-Band Gaussian Noise

In this detection problem the observed waveform is the velope of a narrow-band noise and an additive sine wave amplitude A whose frequency is equal to the center quency of the noise band. The noise is a Gaussian ocess with a zero mean and mean-square value s^2. der these conditions the *pdf* $g_\theta(z)$ is given by [11]

$$z) = \frac{z}{s^2}\exp\left(-\left(\frac{z^2}{2s^2} + \theta\right)\right)I_0\left((z)\left(\frac{2\theta}{s^2}\right)^{1/2}\right), \quad z \geq 0$$
$$= 0, \quad z < 0 \quad (31)$$

ere $I_0(u)$ is the modified Bessel function of the first d, zero order, and $\theta = A^2/2s^2$ is the signal-to-noise wer ratio. The function $b(z)$ is

$$b(z) = z^2/2s^2 - 1, \quad (32)$$

that the locally optimum detector is based on the tistic

$$\frac{1}{n}\sum_{i=1}^{n}((Z_i^2/2s^2) - 1). \quad (33)$$

The Neyman-Pearson detector is based on the statistic

$$\frac{1}{n}\sum_{i=1}^{n}\ln I_0[(Z_i)(2\theta/s^2)^{1/2}], \quad (34)$$

that its structure is much more complicated than that the locally optimum detector. We note that the strucre of the Neyman-Pearson detector depends on θ so that no uniformly optimum detector exists for this problem.

Detection of a Sine Wave of Unknown Phase in Additive Gaussian Noise

The signal in this case is a sine wave of amplitude A, and unknown phase, while the noise is normally distributed with mean zero and variance s^2. Under these conditions the *pdf* $g_\theta(z)$ is given by [12]

$$g_\theta(z) = \frac{1}{\pi s}\int_0^\pi \phi\left(\frac{z}{s} - (2\theta)^{1/2}\cos(u)\right)du \quad (35)$$

where

$$\phi(u) = (2\pi)^{-1/2}\exp(-\tfrac{1}{2}u^2) \quad (36)$$

and $\theta = A^2/2s^2$ is the signal-to-noise power ratio. The function $b(z)$ is

$$b(z) = \tfrac{1}{2}(z^2/s^2 - 1), \quad (37)$$

so that the locally optimum detector is based on the statistic

$$\frac{1}{n}\sum_{i=1}^{n}((Z_i^2/s^2) - 1). \quad (38)$$

The Neyman-Pearson detector is based on the statistic

$$\frac{1}{n}\sum_{i=1}^{n}\left((Z_i^2/2s^2) + \ln\int_0^\pi \phi\left(\frac{Z_i}{s} - (2\theta)^{1/2}\cos u\right)du\right), \quad (39)$$

so that its structure is much more complicated than that of the locally optimum detector.

Bibliography

[1] E. Reich and P. Swerling, "Detection of a sine wave in Gaussian noise," *J. Appl. Phys.*, vol. 24, pp. 289–296; March, 1953.
[2] D. Middleton, "Statistical criteria for the detection of pulsed carriers in noise: I, II," *J. Appl. Phys.*, vol. 24, pp. 371–378, 379–391; April, 1953.
[3] W. W. Peterson, T. G. Birdsall, and W. C. Fox, "The theory of signal detectability," IRE Trans. on Information Theory, no. PGIT-4, pp. 171–212; September, 1954.
[4] E. L. Lehmann, "The power of rank tests," *Ann. Math. Stat.*, vol. 24, pp. 23–43; March, 1953.
[5] H. Cramér, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J., p. 67; 1951.
[6] E. L. Lehmann, "Testing Statistical Hypotheses," John Wiley and Sons, Inc., New York, N. Y., p. 65; 1959.
[7] E. J. G. Pitman, "Lecture Notes on Nonparametric Statistical Inference, Columbia University, New York, N. Y., p. 54; Spring, 1948.
[8] D. A. S. Fraser, "Nonparametric Methods in Statistics," John Wiley and Sons, Inc., New York, N. Y., p. 270; 1957.
[9] J. Capon, "A nonparametric technique for the detection of a constant signal in additive noise," 1959 IRE WESCON Convention Record, pt. 4, pp. 92–103.
[10] J. Capon, "Optimum coincidence procedures for detecting weak signals in noise," 1960 IRE International Convention Record, pt. 4, pp. 154–166.
[11] S. O. Rice, "Mathematical analysis of random noise," in "Noise and Stochastic Processes," N. Wax, Ed., Dover Publications, Inc., New York, N. Y., p. 238; 1954.
[12] S. O. Rice, "Statistical properties of a sine wave plus random noise," *Bell Sys. Tech. J.*, vol. 27, pp. 109–157; January, 1948.
[13] D. Middleton, "An Introduction to Statistical Communication Theory," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 903–906; 1960.
[14] S. Kullback, "Information Theory and Statistics," John Wiley and Sons, Inc., New York, N. Y., pp. 26–28; 1959.

Frequency Differences Between Two Partially Correlated Noise Channels*

JANIS GALEJS†, MEMBER, IRE

Summary—Approximate probability distributions of the difference frequency between two noise channels which contain dissimilar Gaussian, rectangular or triple-tuned RLC band-pass filters are calculated. For noise channels that differ only in time delay, a proportionality between rate of change of instantaneous frequency and the difference frequency is assumed. For dissimilar filters, an approximately equivalent single filter-time delay process is defined. The single filter is determined from the moment averages of the two dissimilar filters, while the equivalent time delay is computed by equating the magnitude of the correlation function in the two processes.

I. Introduction

The phase difference distribution between two Gaussian channels has been examined by several authors under assumptions of stationary or nonstationary noise or with sinusoidal and Rayleigh distributed signals.[1-3] The phase difference distributions characterize performance of phase comparison systems. Similarly, the difference frequency distributions characterize the performance of frequency comparison systems. A simple model of the latter system is provided by two correlated noise channels, the difference frequency of which is applied to an ideal frequency detector. The amplitude distribution of the frequency detector output is the same as the distribution of the instantaneous difference frequency between the two noise channels. Analysis of FM receiver noise and of fading sinusoidal carriers may be quoted among possible applications of the above model. The distribution of noise output changes of an FM receiver between times t and $t' = t + \tau$ may be computed as the difference frequency between two correlated noise channels that differ only in time delay τ. Under the assumption that narrow-band noise exhibits the frequency modulation of a fading carrier, two fading carriers and the associated receiver circuitry exemplify dissimilar noise channels.

Some of the problems in deriving the difference frequency distribution will become apparent from relations between the difference frequency and the noise components of the two channels. The noise output of the nth channel may be represented by

$$V_{on}(t) = X_n(t) \cos \omega_0 t - Y_n(t) \sin \omega_0 t$$
$$= \sqrt{X_n^2(t) + Y_n^2(t)} \cos [\omega_0 t + \phi_n(t)], \qquad (1)$$

where

$$\phi_n(t) = \tan^{-1} \frac{Y_n(t)}{X_n(t)}. \qquad (2)$$

Provided that the noise components $X_n(t)$ and $Y_n(t)$ are defined with respect to the same frequency, ω_0, the phase difference between the two channels becomes

$$\phi_2(t) - \phi_1(t) = \tan^{-1} \frac{Y_2(t)}{X_2(t)} - \tan^{-1} \frac{Y_1(t)}{X_1(t)}. \qquad (3)$$

The frequency difference is simply the time derivative of (3). Differentiation of (3) shows that $\dot{\phi}_2 - \dot{\phi}_1$ depends on 8 variables: X_1, Y_1, X_2, Y_2, $\dot{X}_1$, $\dot{Y}_1$, $\dot{X}_2$, $\dot{Y}_2$. With X_n and Y_n Gaussian, the derivatives $\dot{X}_n$ and $\dot{Y}_n$ are also Gaussian, provided that the autocorrelation functions of $\dot{X}_n$ and $\dot{Y}_n$ exist.[4] The difference frequency $\dot{\phi}_2 - \dot{\phi}_1$ is characterized by an eight-dimensional Gaussian distribution, and the $\dot{\phi}_2 - \dot{\phi}_1$ probability density may be determined for arbitrarily dissimilar noise channels by integrating this Gaussian distribution. Thus, a transformation of $w(X_1, Y_1, X_2, Y_2, \dot{X}_1, \dot{X}_2, \dot{Y}_1, \dot{Y}_2)$ to polar coordinates gives $w(R_1, R_2, \phi_1, \phi_2, \dot{R}_1, \dot{R}_2, \dot{\phi}_1, \dot{\phi}_2)$ which may be reduced to $w(\dot{\phi}_2 - \dot{\phi}_1)$ after seven integrations. The characteristic function method[5] introduces 8 additional integrals. An effort along these two lines undertaken by members of this laboratory has not yielded results readily suited for a numerical evaluation of $w(\dot{\phi}_2 - \dot{\phi}_1)$. In a different approach,[6] the cumulative probability of a derivative is formally related to the joint probability density of the original variable at two time instants t and $t' = t + \tau$, but in our case, evaluation of the density $w[(\phi_2 - \phi_1)_t, (\phi_2 - \phi_1)_{t'}]$ involves the integration of an eight-dimensional Gaussian distribution. Similarly, difficulties may be expected by trying to obtain $w(\dot{\phi}_2 - \dot{\phi}_1)$ from the second order characteristic function of $(\phi_2 - \phi_1)$.

* Received by the PGIT, May 13, 1960; revised manuscript received, August 4, 1960. The research upon which this paper is based was supported by a contract of the Dept. of Defense.

† Appl. Res. Lab., Sylvania Electronic Sys., Waltham, Mass.

[1] V. V. Zvetnov, "Statistical properties of signals and noise in two channel phase systems," *Radiotekhnika (Moscow)*, vol. 12, no. 5, pp. 12–30; 1957.

[2] M. S. Aleksandrov, "Distribution of changes in phase difference for fluctuating random signals and correlated random noise," *Radiotekh. Elektron.*, vol. 5, no. 3, pp. 360–365; 1960.

[3] P. Bello, "Demodulation of a Phase Modulated Noise Carrier," IRE Trans. on Information Theory, vol. IT-7, pp. 19–27; January, 1961.

[4] David Middleton, "An Introduction to Statistical Communication Theory," McGraw-Hill Book Co., Inc., New York, N. Y. sect. 8.1-1; 1960.

[5] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 24, sect. 13.3; 1950.

[6] J. E. Moyal, "Stochastic processes and statistical physics," *J. Roy. Statistical Soc., Ser. B*, vol. 11, pp. 150–210; 1949. See especially p. 179.

The problem may be simplified by imposing restrictions the two noise channels. For almost fully correlated ise channels that differ only in time delay (hence time lays τ are small relative to the reciprocal of channel ndwidth B), one may compute the frequency difference $+ 0.5\tau) - \dot{\phi}(t - 0.5\tau)$ from the phase differences $+ \tau) - \phi(t)$ and $\phi(t) - \phi(t - \tau)$, or from $\phi(t)$, $\dot{\phi}(t)$, d $\ddot{\phi}(t)$. The difference frequency is now expressed in rms of the noise components X_1, Y_1, X_2, Y_2, X_3, Y_3 or terms of noise components X, Y and their derivatives $\dot{Y}$, $\ddot{X}$, $\ddot{Y}$. The distribution of $\dot{\phi}_2 - \dot{\phi}_1$ may be determined om the distribution of only six Gaussian variables.

It is straightforward to obtain the probability density the rate of frequency changes $w(\ddot{\phi})$ from the density $X, Y, \dot{X}, \dot{Y}, \ddot{X}, \ddot{Y}$). For sufficiently small time increents, $w(\dot{\phi}_2 - \dot{\phi}_1)$ is proportional to $w(\ddot{\phi})$ and to τ^{-1}, and e cumulative probability $P(|\dot{\phi}_2 - \dot{\phi}_1| \leq \Delta\dot{\phi}_0)$ may be timated from $P(|\ddot{\phi}| \leq \ddot{\phi}_0)$. This development for dermining $(\dot{\phi}_2 - \dot{\phi}_1)$ distributions of channels that differ ly in time delay is shown in Sections II and III and the sults are summarized in the first part of Section IV.

Additional approximations are required for dissimilar ter channels. An equivalent Gaussian process will be troduced whose output components X_1, Y_1, $\dot{X}_1$, $\dot{Y}_1$ and Y_2, $\dot{X}_2$, $\dot{Y}_2$ separated in time by τ approximate the itput components of two dissimilar filters X_1, Y_1, $\dot{X}_1$, X_2, Y_2, $\dot{X}_2$, $\dot{Y}_2$. This is achieved by attempting to atch the second moments of the two eight-dimensional aussian distributions and by defining the delay time τ the equivalent Gaussian process by equating the agnitudes of the two correlation functions as indicated Section IV. Once the approximately equivalent Gaussian ocess (including τ) is defined, the cumulative probility $P[(\dot{\phi}_2 - \dot{\phi}_1) < \Delta\,\dot{\phi}_0]$ is computed as it would be r channels that differ only in time delay. It is not shown w close the difference frequency distribution of the proximate model comes to the actual difference freiency distribution between the two dissimilar channels, it one can define equivalent channels which will give o large or too small spreads of difference frequencies. ne actual difference frequency distribution appears to between the above bounds and may be expected to be ose to the approximate value, as indicated in Appendix I.

II. Joint Probability Density of $\dot{\phi}$ and $\ddot{\phi}$ and Its Integrals

The probability density of the rate of frequency change) and the cumulative probability $P(|\ddot{\phi}| < \ddot{\phi}_0)$ are comited from the joint probability density $w(\dot{\phi}, \ddot{\phi})$. The joint obability density $w(\dot{\phi}, \ddot{\phi})$ is obtained by integrating a -dimensional Gaussian distribution of the variables Y, $\dot{X}$, $\dot{Y}$, $\ddot{X}$ and $\ddot{Y}$. Transforming (3.8–4) of Rice[7] to polar coordinates, one obtains $w(R, \dot{R}, \ddot{R}, \phi, \dot{\phi}, \ddot{\phi})$. Now

$$w(\dot{\phi}, \ddot{\phi}) = \int_0^{2\pi}\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(R, \dot{R}, \ddot{R}, \phi, \dot{\phi}, \ddot{\phi})\, d\phi\, dR\, d\dot{R}\, d\ddot{R}. \quad (4)$$

The R integration is elementary. The $\ddot{R}$ integration may be carried out after changing the variable of integration to $(\ddot{R} - R\dot{\phi}^2)$. There are no difficulties with R and ϕ integrations. The integrations result in

$$w(\dot{\phi}, \ddot{\phi}) = \frac{(\rho_2 - \mu_1^2)^{1.5}F^{1.5}}{\pi[(\rho_2 - 2\mu_1\dot{\phi} + \dot{\phi}^2)F + (\rho_2 - \mu_1^2)\ddot{\phi}^2]^2}\,; \quad (5)$$

where

$$F = (\rho_4 - \rho_2^2) + 4(\mu_1\rho_2 - \mu_3)\dot{\phi} + 4(\rho_2 - \mu_1^2)\dot{\phi}^2, \quad (6)$$

$$\mu_1\sigma^2 = \mu'(0) = 2\pi\int_0^{\infty} g(f)(f - f_0)\, df, \quad (7)$$

$$\rho_2\sigma^2 = -\rho''(0) = (2\pi)^2\int_0^{\infty} g(f)(f - f_0)^2\, df, \quad (8)$$

$$\mu_3\sigma^3 = -\mu'''(0) = (2\pi)^3\int_0^{\infty} g(f)(f - f_0)^3\, df, \quad (9)$$

$$\rho_4\sigma^2 = \rho^{\text{iv}}(0) = (2\pi)^4\int_0^{\infty} g(f)(f - f_0)^4\, df, \quad (10)$$

$$\sigma^2 = \int_0^{\infty} g(f)\, df, \quad (11)$$

and where $g(f)$ is the noise power spectrum and f_0 designates the frequency with respect to which the noise components X and Y are defined. The moments (7) to (11) may be derived from (61) of Appendix I. The above moments expressed as time integrals are shown in (49), (55), (57), (59), and (60) of the same Appendix. Moments for specific filter characteristics are listed in Appendix II.

The probability density $w(\dot{\phi}, \ddot{\phi})$ may be integrated to obtain $w(\dot{\phi})$. Thus

$$w(\dot{\phi}) = \frac{\rho_2 - \mu_1^2}{2[\rho_2 - 2\mu_1\dot{\phi} + \dot{\phi}^2]^{1.5}}. \quad (12)$$

Although it has been possible to obtain an explicit expression for $w(\ddot{\phi})$ with symmetrical spectra,[8] an attempt to evaluate the integral with $\mu_1 \neq 0$ and $\mu_3 \neq 0$ was successful only for large values of $\ddot{\phi}$. Hence,

$$w(\ddot{\phi}) \approx 1/\pi\,\frac{\rho_2 - \mu_1^2}{\ddot{\phi}^2}. \quad (13)$$

For smaller arguments $\ddot{\phi}$, $w(\ddot{\phi})$ is obtained numerically. The cumulative probability $P(|\ddot{\phi}| < \ddot{\phi}_0)$ is computed by numerically integrating $w(\ddot{\phi})$. The cumulative probability has been plotted in Fig. 1 for Gaussian, rectangular and

[7] S. O. Rice, "Mathematical analysis of random noise," *Bell s. Tech. J.*, vol. 23, pp. 282–332, July, 1944; vol. 24, pp. 46–156; nuary, 1945.

[8] Introducing a variable $v = 4\phi^2\rho_2^2[4\phi^2\rho_2^2 + (\rho_4 - \rho_2^2)^2]^{-1}$ and setting $\mu_1 = \mu_3 = 0$, $w(\ddot{\phi})$ may be computed with the aid of integrals 212.10 and 212.11 of W. Groebner and N. Hofreiter, "Integraltafel," Springer Verlag, Vienna, Austria, pt. 1, 1949.

triple-tuned RLC band-pass[9] characteristics with $\mu_1 = \mu_3 = 0$. It is seen that for a given rate of frequency change[10]

$$\ddot{\phi}_0 = k(\pi B)^2, \tag{14}$$

the cumulative probability $P(|\ddot{\phi}| \leq \ddot{\phi}_0)$ is smallest for the triple RLC filter. A filter with the most gradual cutoff exhibits the largest amount of high-frequency output components. Its instantaneous output frequency will be extended over a wider frequency range ($w(\dot{\phi})$ is a broader distribution). High values of $\ddot{\phi}$ are more likely in a gradual cutoff filter (also a broad $w(\ddot{\phi})$ distribution), which is the cause of the lower $P(|\ddot{\phi}| \leq \ddot{\phi}_0)$ values for $\ddot{\phi}_0/(\pi B)^2 =$ constant.

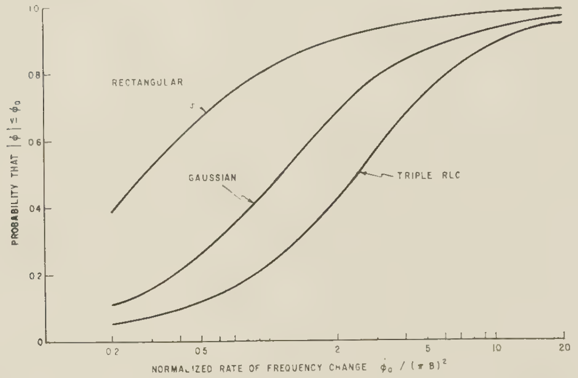


Fig. 1—Probability of the rate of frequency changes.

III. Relations Between Probabilities Involving Increments and Derivatives of a Variable

A simple proportionality that exists at sufficiently small time increments τ between the probability density for a change of a variable Δx and the probability density of its derivative $\dot{x}$ is the basis of the approximate method described in this paper. From

$$\Delta x = x_2 - x_1 = \tau\dot{x} + 0(\tau^2), \tag{15}$$

it follows that the probability densities for Δx and $\dot{x}$ are related by

$$w(\Delta x, \tau) \approx \frac{w(\dot{x})}{\tau}. \tag{16}$$

Furthermore, integrating Δx from $-\Delta x_0$ to $+\Delta x_0$ shows that the cumulative probabilities are related by

$$P(|\Delta x| \leq \Delta x_0, \tau) = P\left(|\dot{x}| \leq \frac{\Delta x_0}{\tau}\right). \tag{17}$$

An example of the relation between the incremental and derivative probability may be provided by deriving (12) from the probability density $w(\Delta\phi, \tau)$. For strong noise correlation (τ small), $w(\Delta\phi, \tau)$ may be approximated by[11]

$$w(\Delta\phi, \tau) \approx \frac{(\sigma^4 - \rho^2 - \mu^2)\beta}{2\sigma^4(1 - \beta^2)^{1.5}}, \tag{18}$$

where

$$\beta = (\rho \cos \Delta\phi + \mu \sin \Delta\phi)\sigma^{-2} \approx 1, \tag{19}$$

and where σ^2, ρ and μ are defined by (11) and (61). Expanding ρ and μ in powers of τ and using the small angle approximations of sine and cosine functions, one may obtain (12) by using (16).

The probabilities of the rate of frequency change and of frequency increments are related by (16) and (17). Thus,

$$w(\Delta\dot{\phi}, \tau) = \frac{1}{\tau} w(\ddot{\phi}), \tag{20}$$

$$P(|\Delta\dot{\phi}| \leq \Delta\dot{\phi}_0, \tau) = P\left(|\ddot{\phi}| \leq \frac{\Delta\dot{\phi}_0}{\tau}\right). \tag{21}$$

Let the cumulative probability

$$P[|\ddot{\phi}| \leq \ddot{\phi}_0] = m \tag{22}$$

be given. Rewriting (14) and comparing the left-hand side of (22) with the right-hand side of (21), one gets

$$\ddot{\phi}_0 = k(\pi B)^2 = \Delta\dot{\phi}_0/\tau, \tag{23}$$

where the bandwidth B is defined above.[10] Substituting (23) in (21) and equating (21) and (22) gives

$$P[|\Delta\dot{\phi}| < k(\pi B\tau)\pi B, \tau] = m. \tag{24}$$

IV. Distribution of the Difference Frequency

A. Identical Band-Pass Characteristics

The simplest case to consider involves identical band-pass filters in two noise channels and a time delay in one of the channels as indicated in Fig. 2. With the cumulative probability $P(|\ddot{\phi}| \leq \ddot{\phi}_0) = m$ plotted in Fig. 1 and

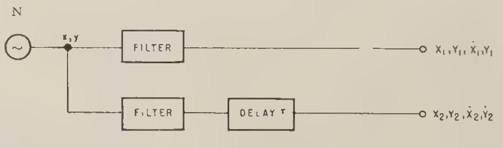


Fig. 2—Block diagrams of a two-channel system with identical filters and a time delay in one channel.

[9] A synchronously triple-tuned RLC filter is the simplest physically realizable band-pass filter for which $w(\ddot{\phi})$ may be defined. The autocorrelation function of the output noise from a single-tuned RLC filter has a discontinuous first derivative at $\tau = 0$. For a double-tuned RLC filter, the third derivative of the autocorrelation function is discontinuous. The parameters ρ_2 and ρ_4 that are required for specifying $w(\dot{\phi}, \ddot{\phi})$ are infinite for a single-tuned RLC filter. For a double-tuned filter, ρ_4 cannot be defined.

[10] $\ddot{\phi}$ is normalized with respect to the bandwidth B squared of the three types of filters considered. As defined in Appendix II, B in cps is the bandwidth of the rectangular filter, the $e^{-0.5}$ bandwidth of the Gaussian filter, and the half-power bandwidth of the triple-tuned RLC filter.

[11] W. B. Davenport and W. L. Root, "Random Signals and Noise," McGraw-Hill Book Co., Inc., New York, N. Y., 1958. Only the term $\beta\pi$ of the numerator in Eq. (8.106) is significant for $\beta \approx 1$.

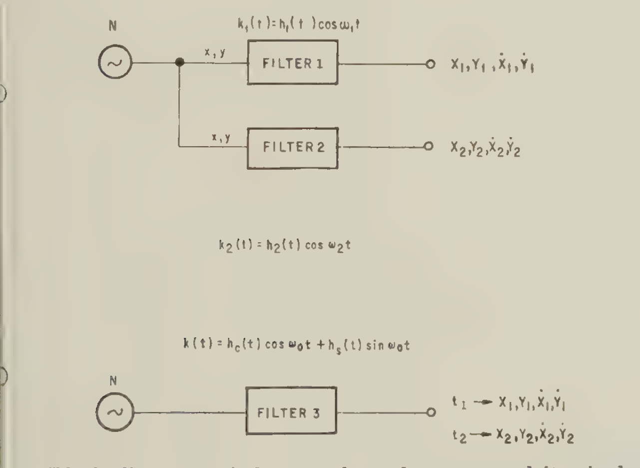
time delay τ given, one may compute $\Delta\phi_0$ in $P(|\Delta\phi| \leq$ $, \tau) = m$ from (23). This computational procedure will nost accurate for small values of τ or for almost fully elated noise in the outputs of the two channels.

Dissimilar Band-Pass Filters, Aligned Center Fre- uencies

he system involving dissimilar band-pass filters may analyzed by means of the relations developed for ilar filters and for a time delay. The system depicted ig. 3(a) will be approximated by a system with similar d-pass filters as shown in Fig. 2, or by a system in- ving a single filter as shown in Fig. 3(b). It is per- sible to substitute the system of Fig. 3(b) for the one ig. 3(a) if the second moments of the Gaussian distri- ion that characterize the system outputs are the same both cases. However, a comparison of the moments ed in Appendix I indicates that it is not possible to tch all of the moments of Fig. 3(a) with moments of . 3(b). Introducing the complex notation

$$Z_n = X_n + jY_n, \tag{25}$$

ments like $\overline{Z_m^* \dot{Z}_n}$, $\overline{Z_n^* \dot{Z}_n}$, $\overline{\dot{Z}_n^* Z_n}$ (also $\overline{\dot{Z}_n^* \ddot{Z}_n}$ and $\overline{\ddot{Z}_n^* \dot{Z}_n}$) Fig. 3(a) will have two distinct values as long as the filters in Fig. 3(a) are dissimilar $[h_1(x) \neq h_2(x)]$. The responding expression of Fig. 3(b) have a single value, filter outputs in Fig. 3(b) may only approximate the puts of Fig. 3(a).



3—Block diagrams of the two-channel system and its single-channel approximation.

he approximate procedure for computing the prob- lity distribution of the difference frequency involves

) The specification of the moments μ_r, μ_3, ρ_2, and ρ_4 of (7) to (10) for computing $w(\phi, \ddot{\phi})$ and $P(|\ddot{\phi}| < \ddot{\phi}_0)$. The above moments, which are also listed in Appendix I-B, must be determined from moments listed in Appendix I-A.

) The determination of an equivalent sampling interval $\tau = t_2 - t_1$ in Fig. 3(b) in order to relate $P(|\ddot{\phi}| < \dot{\phi}_0)$ to $P(|\Delta\phi| < \Delta\phi_0)$ as in (21).

As long as there are two different sets of moments corresponding to μ_1, μ_3, ρ_2 and ρ_4 in Fig. 3(a), their average may be used for computing moments of Fig. 3(b). Thus,

$$\mu_1 \sigma^2 = \mu_b'(0) = \frac{\sigma^2}{2}(\mu_{1a(1)} + \mu_{1a(2)}), \tag{26}$$

$$\rho_2 \sigma^2 = -\rho_b''(0) = \frac{\sigma^2}{2}(\rho_{2a(1)} + \rho_{2a(2)}), \tag{27}$$

$$\mu_3 \sigma^2 = -\mu_b'''(0) = \frac{\sigma^2}{2}(\mu_{3a(1)} + \mu_{3a(2)}), \tag{28}$$

$$\rho_4 \sigma^2 = \rho_b^{\text{iv}}(0) = \frac{\sigma^2}{2}(\rho_{4a(1)} + \rho_{4a(2)}). \tag{29}$$

This choice of moments for Fig. 3(b) minimizes the mean-square difference between the moments of the Gaussian distributions characterizing Figs. 3(a) and 3(b); the best approximation in the least-squares sense, to two different quantities, is their average.

It is desired to determine the time interval $\tau = t_2 - t_1$ of Fig. 3(b) in such a way that the probability distribution of the difference frequency in Fig 3(b) becomes the same as in Fig. 3(a). This could also be accomplished by selecting τ such that all the moments of Fig. 3(b) involving $\tau \neq 0$ become approximately equal to the corresponding moments of Fig. 3(a). This provides a total of 3 complex equations [(61)–(63)] for determining a single real variable τ. Because of the approximations made, it will not be possible to satisfy all of the equations. Rather than attempting a least-squares fit of the equations or another method for approximately satisfying all of the constraints, τ will be determined from moments not involving derivatives by

$$\rho_a^2 + \mu_a^2 = \rho_b^2 + \mu_b^2, \tag{30}$$

where ρ_a, μ_a, ρ_b and μ_b are given by (48) and (61). For the constant mean-square noise output of filters in Fig. 3(c) and Fig. 3(b) ($\sigma_1^2 = \sigma_2^2 = \sigma^2$), this specifies identical probability distributions of the phase difference.[12] With (30) satisfied, it is possible to obtain various probability distributions of difference frequency. A heuristic argument[13] may be used to show that the moments (26)–(29), used in conjunction with τ from (30), give a probability distribution of difference frequency which lies between two limiting frequency distributions. The limiting distributions appear to have a wider and narrower spread of difference frequencies than the actual distribution of Fig. 3(a). Although it has not been determined how close this approximation comes to the actual differene frequency distribution, the two limiting distributions may be so close that a more accurate error calculation is not warranted. The computation of τ is made from (30) after

[12] It follows from (12) that $w(\phi - \mu_1) = [(1 - \beta_0^2)/\tau^2] \cdot [(1 - \beta_0^2)/\tau^2 + (\phi - \mu_1)^2]^{-1.5}$ where $\beta_0^2 = (\rho^2 + \mu^2)/\sigma^4 \approx 1 - \tau^2(\rho_2 - \mu_1^2)$, and is related by (16) to $w(\Delta\phi - \mu_1\tau)$. The probability distribution of the phase difference $\Delta\phi$ about its average $\mu_1\tau \approx \mu/\sigma^2$ is thus determined by β_0^2 which has been also shown by Bello, *op. cit.*

[13] See Appendix III.

expanding the moments ρ_b and μ_b. Thus,

$$\rho_b/\sigma^2 \approx 1 - \frac{\tau^2}{2}\rho_2 + \frac{\tau^4}{24}\rho_4, \tag{31}$$

$$\mu_b/\sigma^2 \approx \tau\mu_1 - \frac{\tau^3}{6}\mu_3. \tag{32}$$

Substituting (31) and (32) in (30) and neglecting powers of τ higher than τ^4, one gets

$$(\rho_a^2 + \mu_a^2)/\sigma^4 \approx 1 - \tau^2(\rho_2 - \mu_1^2) + \tau^4\left(\frac{\rho_2^2}{4} + \frac{\rho_4}{12} - \frac{\mu_1\mu_3}{3}\right). \tag{33}$$

Solving (33) for τ^2, one has

$$\tau^2 = \frac{\rho_2 - \mu_1^2 - \sqrt{\rho_2^2\left(\frac{\rho_a^2+\mu_a^2}{\sigma^4}\right) - \left(1 - \frac{\rho_a^2+\mu_a^2}{\sigma^4}\right)\left(\frac{\rho_4}{3} - \frac{4}{3}\mu_1\mu_3\right) - 2\mu_1^2\rho_2 + \mu_1^4}}{\left(\frac{\rho_2^2}{2} + \frac{\rho_4}{6} - \frac{2\mu_1\mu_3}{3}\right)}. \tag{34}$$

A less accurate value of τ^2 is obtained by neglecting the τ^4 term in (33). This gives

$$\tau^2 \approx \frac{1 - \frac{\rho_a^2 + \mu_a^2}{\sigma^4}}{\rho_2 - \mu_1^2}. \tag{35}$$

C. Filters with Misaligned Center Frequencies

The previous calculating procedure must be modified if the two filters of Fig. 3(a) differ in center frequencies. For symmetrical band-pass filters, the probability density of the instantaneous frequency deviations $w(\dot\phi)$ in (12) is symmetrical about the filter center frequency ω_i, and the average value of the instantaneous frequency $\omega_i + \dot\phi$ is simply ω_i. The average value of the frequency difference between the two filters in Fig. 3(a) is the difference between these center frequencies, $\omega_1 - \omega_2$. The difference frequency of Fig. 3(a) was estimated from the rate of frequency change of $\ddot\phi$ in Fig. 3(b). The average value of $\ddot\phi$ should satisfy

$$\bar{\ddot\phi} = \overline{\Delta\dot\phi}/\tau = (\omega_1 - \omega_2)/\tau, \tag{36}$$

according to the earlier development. The probability density $w(\ddot\phi)$ computed from (5) is symmetrical about $\ddot\phi = 0$. $\bar{\ddot\phi} \neq 0$ and may satisfy (36) only if the center frequency of the filter in Fig. 3(b) is swept. Such a frequency sweep may be accounted for by changing the variable $\ddot\phi$ to $(\ddot\phi - \bar{\ddot\phi})$ in (5). The noise components X_1Y_1 and X_2Y_2 can be defined with respect to the instantaneous center frequency of the filter at times τ_1 and τ_2, respectively. The moments μ_1, ρ_2, μ_3 and ρ_4 are computed for the individual filters and their average is used for the computation of $w(\ddot\phi)$, similar to that indicated in (26)–(29). The equivalent delay time τ is computed as in (30), or in (34) and (35).

D. Additive Noise and Nonstationary Channels

Additive noise and nonstationary channels can be handled by methods described earlier.[3]

Additive noise increases the total noise-power output of a channel without increasing the correlated noise portion. The normalized correlation coefficient of the channels is decreased and $(\rho_a^2 + \mu_a^2)/(\sigma_1^2\sigma_2^2)$ should be replaced by

$$\frac{\rho_a^2 + \mu_a^2}{(\sigma_1^2 + \sigma_{1(\text{add})}^2)(\sigma_2^2 + \sigma_{2(\text{add})}^2)},$$

where σ_n and $\sigma_{n(\text{add})}$ are the RMS values of the original and the additive noise of the nth channel, respectively. Increasing amounts of additive noise decrease the left-hand side of (30), which results in increased equivalent time delays τ from (34) or (35).

In the quasi-stationary approximation, time-varying channels can be analyzed by considering the time variation of the instantaneous channel parameters. Thus, time-varying doppler shifts or a frequency modulation is accounted for by introducing appropriately time-varying filter center frequencies. The quasi-stationary approximation becomes inaccurate for rapid changes of channel parameters.[3,14]

V. Discussion

The problem of determining difference frequency distributions between noise channels that differ only in time delay is fairly straightforward. This was shown in Section IV-A and will not be discussed in more detail.

The approximate method for analyzing channels with dissimilar filters involved matching of second moments between the two-filter system to be analyzed and the equivalent single-filter time-delay system. This approximation will be compared with several other single-filter time-delay systems. A number of such equivalent systems for approximating the difference frequency between the triple RLC filters of relative bandwidth difference $\delta = (B_2 - B_1)/B_1 = 0.05$ and of relative center frequency misalignment $\Omega = (f_1 - f_2)/B_3 = 0.2$, where $B_3 = 0.5(B_1 + B_2)$, is depicted in Fig. 4 along with the corresponding 50 per cent and 90 per cent confidence intervals of the difference frequency $\Delta\dot\phi_1$ and $\Delta\dot\phi_2$. The system of Fig. 4(b) provides the best moment match with Fig. 4(a)

[14] E. J. Baghdady, "Theory of low distortion reproduction of FM signals in linear systems," IRE Trans. on Circuit Theory, vol. CT-5, pp. 202–214; September, 1958.

ne filters of Fig. 4(c) are of the same type as in Fig. 4(a), ut of intermediate bandwidth and center frequencies. ne second moments of Fig. 4(c) do not depend on the lative bandwidth difference δ of the filters of Fig. 4(a) ne system of Fig. 4(c) is most easily handled analytically nce only the equivalent time delay τ is affected by δ nd Ω. The systems of Figs. 4(d) or 4(e) and of Figs. f) or 4(g) provide difference frequency distributions at are narrower and wider than the unknown distri- ition of Fig. 4(a). Figs. 4(e) and 4(g) illustrate noise ectra that are skew about f_0. The numerical difference equency figures indicate a negligible change between igs. 4(b) and 4(c). This change will be decreased even rther with more accurately aligned filters.[15] The upper and lower bounds of difference frequencies[16] are separated by δ, which is 5 per cent in the examples of Fig. 4.

The previous example tends to justify the use of the equivalent diagram in Fig. 4(c). This diagram will be used in comparisons between Gaussian, rectangular, and triple-tuned RLC filters. Table I summarizes the 50 per cent and 90 per cent confidence intervals of the difference frequency $\Delta\phi_1$ and $\Delta\phi_2$ that are normalized with respect to the average bandwidth B_3. It is seen that rectangular filters exhibit the largest $\Delta\phi_1$ and $\Delta\phi_2$ values for close-filter tolerances (δ and Ω small). Although for $P(|\ddot{\phi}| \leq \ddot{\phi}_0) =$ constant in Fig. 1, the rectangular filter exhibits the smallest $\ddot{\phi}/(\pi B)^2$ values, $\Delta\phi$ is proportional to both τ and $\ddot{\phi}$ from (23). The ratio of the equivalent time delay between the rectangular filter and the other filters is

[15] The moments ρ_2 and ρ_4 are larger in Fig. 4(c) by a factor $+ x$), where x is proportional to (δ^2). Increased moments de- ease $w(\phi, \ddot{\phi})$ of (5) and increase the value $\ddot{\phi}_0$ for $P(|\ddot{\phi}| \leq \ddot{\phi}_0) =$ nstant. This change in $\ddot{\phi}_0$ is decreased as δ decreases.

[16] Change in difference frequencies between Fig. 4(d) or 4(e) and Figs. 4(f) or 4(g).

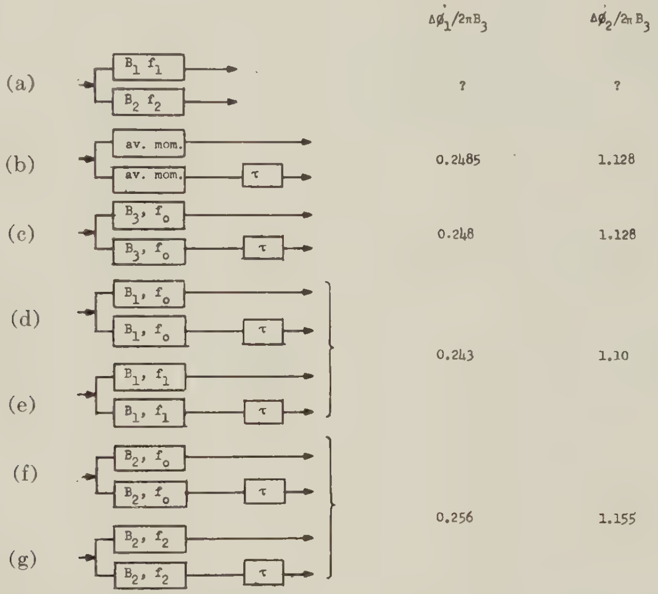


Fig. 4—Equivalent diagrams for computing difference frequency between two dissimilar triple-tuned RLC band-pass filters. In Fig. 4 (c)–(g), $X_1 Y_1 X_2 Y_2$ are defined with respect to f_0.

$f_0 = 0.5(f_1 + f_2)$ $B_3 = 0.5(B_1 + B_2)$
$P(|\Delta\phi| \leq \Delta\phi_1) = 0.5$ $P(|\Delta\phi| \leq \Delta\phi_2) = 0.9$
$\delta = (B_2 - B_1)/B_1 = 0.05$ $\Omega = (f_1 - f_2)/B_3 = 0.2$

TABLE I
Difference Frequency Between Two Noise Channels with Gaussian, Rectangular and Triple-Tuned RLC Band-Pass Filters

		Gaussian		Rectangular		Triple RLC	
δ	Ω	$\Delta\phi_1/2\pi B_3$	$\Delta\phi_2/2\pi B_3$	$\Delta\phi_1/2\pi B_3$	$\Delta\phi_2/2\pi B_3$	$\Delta\phi_1/2\pi B_3$	$\Delta\phi_2/2\pi B_3$
0.01	0.01	0.0056	0.04	0.035	0.24	0.017	0.077
	0.05	0.028	0.16	0.0775	0.54	0.061	0.278
	0.20	0.011	0.65	0.155	1.08	0.24	1.1
0.05	0.01	0.02	0.12	0.055	0.38	0.061	0.278
	0.05	0.034	0.2	0.0775	0.54	0.085	0.386
	0.20	0.1135	0.66	0.155	1.08	0.248	1.13

Note: $P(|\Delta\phi| \leq \Delta\phi_1) = 0.5$ $P(|\Delta\phi| \leq \Delta\phi_2) = 0.9$
$\delta = (B_2 - B_1)/B_1$ $\Omega = (f_1 - f_2)/B_3$
$B_3 = 0.5(B_1 + B_2)$
The bandwidths B_n of the different filters are defined as in footnote 10 or Appendix II.

largest for small δ or Ω,[17] which causes the larger $\Delta\phi_i$ values in Table I. For larger δ or Ω, the ratio of the equivalent time delays between the rectangular and other filters is decreased and the effect of the larger $\ddot{\phi}/(\pi B)^2$ values of the triple RLC filter tends to predominate. For larger alignment errors δ or Ω, the triple RLC filter exhibits larger $\Delta\phi_i$ values than the sharper cutoff rectangular or Gaussian filters. The data of Table I can be readily extended to other values of δ and Ω by means of the τ relations given in Appendix II and to probabilities $P \neq 0.5, 0.9$, with the aid of Fig. 1.

The above examples are indicative of channel tolerances required for achieving a prescribed level of difference frequency fluctuations. It should be remembered that the accuracy of the results depends on the degree of noise correlation in the two channels. For a high degree of noise correlation or for small dissimilarities in the filter characteristics, the proportionality between the corresponding probability densities is almost exact. Also, the approximation of two different sets of second moments of a Gaussian distribution, by their averages, appears to introduce a smaller error (the per cent difference between the upper and lower bounds of the difference frequency is decreased) under similar conditions.

Appendix I
Moments of the Gaussian Distributions Characterizing Two Noise Channels

A. Dissimilar Filters in the Two Channels

Complex notation[18,19] is introduced in this Appendix in order to condense the derivation of the various second moments. The filter inputs and outputs of Fig. 3(a) may be represented as[20]

$$V_i = x \cos \omega_0 t - y \sin \omega_0 t = Re\,(ze^{j\omega_0 t}), \tag{37}$$

$$V_{on} = X_n \cos \omega_0 t - Y_n \sin \omega_0 t = Re\,(Z_n e^{j\omega_0 t}). \tag{38}$$

The impulse response of the nth filter is

$$k_n(t) = Re\,[h_n(t)e^{j\Delta_n t}e^{j\omega_0 t}], \tag{39}$$

where

$$\Delta_n = \omega_n - \omega_0, \tag{40}$$

and where $h_n(t)$ is real if the filter is symmetrical with respect to its center frequency ω_n. The complex envelopes of the filter outputs

$$Z_n = X_n + jY_n \tag{41}$$

are related to the complex envelopes of the filter inputs

$$z = x + jy, \tag{42}$$

and to the complex envelope of the filter impulse response $h_n e^{j\Delta_n t}$ by

$$Z_n(t) = 0.5 \int z(t-u)h_n(u)e^{j\Delta_n u}\, du. \tag{43}$$

Furthermore, with $h_n(t)$ real,

$$\overline{Z_1^* Z_2} = 0.25 \iint h_1(u)h_2(v)\overline{z^*(u)z(v)}e^{-j(\Delta_1 u - \Delta_2 v)}\, du\, dv. \tag{44}$$

Assuming that

$$\overline{x(u)y(v)} = \overline{x(v)y(u)} = 0, \tag{45}$$

it follows that

$$\begin{aligned}\overline{z^*(u)z(v)} &= 2\overline{x(u)x(v)} \\ &= 2\int_0^\infty g(f)\cos[(\omega-\omega_0)(u-v)]\, df.\end{aligned} \tag{46}$$

For input noise N of constant spectral power density $4\,w_0$, (46) simplifies to

$$\overline{z^*(u)z(v)} = 8w_0\,\delta(u-v). \tag{47}$$

Substituting (47) in (44), one has

$$\rho_a + j\mu_a = 0.5\overline{Z_1^* Z_2} = w_0 \int h_1(u)h_2(u)e^{-j(\Delta_1-\Delta_2)u}\, du; \tag{48}$$

further,

$$\sigma_n^2 = 0.5\overline{Z_n^* Z_n} = w_0 \int h_n^2(u)\, du. \tag{49}$$

Changing the variable of integration in (43) to

$$w = t - u \tag{50}$$

gives

$$Z_n(t) = 0.5 \int_{-\infty}^{w_1} z(w)h_n(t-w)e^{j\Delta_n(t-w)}\, dw, \tag{51}$$

where $w_1 = t$ for physically realizable filters and where $w_1 = \infty$ for physically nonrealizable filters which exhibit a nonzero impulse response for negative times. The differentiation of (51) with respect to t may be carried out under the integral sign, provided that $h_n(0) = 0$ for the realizable filters. Changing back to the u variable gives

$$\dot{Z}_n(t) = 0.5 \int z(t-u)[\dot{h}_n(u) + j\,\Delta_n h_n(u)]e^{j\Delta_n u}\, du. \tag{52}$$

The moments involving derivatives may be computed from (43), (47) and (52) as follows:

[17] $\tau \sim \sqrt{|\delta|}$ or $\sqrt{|\Omega|}$ in (97) for rectangular filter, while $\tau \sim \sqrt{K\delta^2 + \Omega^2}$ for the other filter types in (86) and (109).

[18] R. Arens, "Complex processes for envelopes of normal noise," IRE Trans. on Information Theory, vol. IT-3, pp. 204–207; September, 1957.

[19] J. Dugundji, "Envelopes and pre-envelopes of real waveforms," IRE Trans. on Information Theory, vol. IT-4, pp. 53–57, March, 1958.

[20] The noise components x, y and X, Y may also be defined with respect to the center frequencies of the two filters ω_n. This will make the second moments involving noise components of both filters time varying. However, sums of squared moments like $(\overline{X_1X_2})^2 + (\overline{Y_1Y_2})^2$ or corresponding expressions involving derivatives of the noise components remain constant and are not affected by the various choices of center frequencies.

$$_n + j\dot{\mu}_{an} = 0.5\overline{Z_m^* \dot{Z}_n}$$

$$= w_0 \int h_m(u)[\dot{h}_n(u) + j\,\Delta_n h_n(u)]e^{-j(\Delta_m - \Delta_n)u}\,du, \tag{53}$$

$$+ j\ddot{\mu}_a = 0.5\overline{\dot{Z}_1^* \dot{Z}_2} = w_0 \int [\dot{h}_1(u) - j\,\Delta_1 h_1(u)]$$

$$\cdot[\dot{h}_2(u) + j\,\Delta_2 h_2(u)]e^{-j(\Delta_1 - \Delta_2)u}\,du. \tag{54}$$

ith $m = n$ in (53),

$$_n\mu_{1a(n)} = 0.5\overline{Z_n^* \dot{Z}_n}/j = w_0\,\Delta_n \int h_n^2(u)\,du = \Delta_n \sigma_n^2 \tag{55}$$

nce

$$\int h_n(u)\dot{h}_n(u)\,du = -j \int \omega\,|\,H_n(j\omega)\,|^2\,d\omega = 0. \tag{56}$$

ith $m = n$ in (54),

$$_n\rho_{2a(n)} = 0.5\overline{\dot{Z}_n^* \dot{Z}_n} = w_0 \int [\dot{h}_n^2(u) + \Delta_n^2 h_n^2(u)]\,du. \tag{57}$$

ith

$$_n(t) = 0.5 \int z(t-u)[\ddot{h}_n(u) + 2j\,\Delta_n \dot{h}_n(u)$$

$$- \Delta_n^2 h_n(u)]\cdot e^{j\Delta_n u}\,du, \tag{58}$$

ne additional moments for specifying $w(X, \dot{X}, \ddot{X}, Y,$ $, \ddot{Y})$ are

$$_n\mu_{3a(n)} = \frac{1}{2j}\overline{\ddot{Z}_n^* \dot{Z}_n} = w_0\,\Delta_n \int [2\dot{h}_n^2(u) - \ddot{h}_n(u)h_n(u)]\,du$$

$$+ w_0\,\Delta_n^3 \int h_n^2(u)\,du, \tag{59}$$

$$_n\rho_{4a(n)} = \tfrac{1}{2}\overline{\ddot{Z}_n^* \ddot{Z}_n}$$

$$= w_0 \int \{[\ddot{h}_n(u) - \Delta_n^2 h_n(u)]^2 + [2\,\Delta_n \dot{h}_n(u)]^2\}\,du. \tag{60}$$

. Identical Filters with Time Delay in One Channel

The second moments of the filter output in Fig. 2 or ig. 3(b) are as follows:

$$\rho_b + j\mu_b = 0.5\overline{Z_1^* Z_2} = \int_0^\infty g(f)e^{j(\omega - \omega_0)\tau}\,df, \tag{61}$$

here Z_n is defined by (41) and where $g(f)$ is the power pectrum of the filter ouput. The moments involving erivatives are[21]

$$0.5\overline{Z_1^* \dot{Z}_2} = -0.5\overline{\dot{Z}_1^* Z_2} = \rho_b' + j\mu_b', \tag{62}$$

$$0.5\overline{\dot{Z}_1^* \dot{Z}_2} = -(\rho_b'' + j\mu_b''), \tag{63}$$

$$\sigma^2 \rho_{2b} = 0.5\overline{\dot{Z}_n^* \dot{Z}_n} = -\rho_b''(0), \tag{64}$$

$$\sigma^2 \mu_{1b} = 0.5\overline{Z_n^* \dot{Z}_n}/j = \mu_b'(0). \tag{65}$$

The additional relations for specifying $w(X, \dot{X}, \ddot{X}, Y, \dot{Y}, \ddot{Y})$ are

$$\sigma^2 \mu_{3b} = 0.5\overline{\dot{Z}_n^* \ddot{Z}_n}/j = -\mu_b'''(0), \tag{66}$$

$$\sigma^2 \rho_{4b} = 0.5\overline{\ddot{Z}_n^* \ddot{Z}_n} = \rho_b^{\mathrm{iv}}(0). \tag{67}$$

In the above equations, the dots and primes denote differentiation with respect to t and τ respectively.

Appendix II
Parameters of Specific Filters

A. Gaussian Filter

A Gaussian filter of $e^{-0.5}$ power, bandwidth B_n(cps) has a transfer characteristic

$$H_n(j\omega) = A_n \exp[-(f \pm f_n)^2/B_n^2], \tag{68}$$

where the plus and minus signs refer to negative and positive frequencies respectively. The corresponding impulse response envelope is

$$h_n(t) = 2\sqrt{\pi}\,A_n B_n \exp[-(\pi B_n t)^2]. \tag{69}$$

Computing the moments of individual channels in Fig. 3(a) by (7)–(10) or by (55), (57), (59) and (60),

$$\mu_{1a(n)} = 2\pi(f_n - f_0), \tag{70}$$

$$\rho_{2a(n)} = \pi^2[B_n^2 + 4(f_n - f_0)^2], \tag{71}$$

$$\mu_{3a(n)} = \pi^3(f_n - f_0)[6B_n^2 + 8(f_n - f_0)^2], \tag{72}$$

$$\rho_{4a(n)} = \pi^4[3B_n^4 + 24(f_n - f_0)^2 B_n^2]; \tag{73}$$

where difference frequency terms in powers above the second have been neglected. The moment averages are computed with a reference bandwidth

$$B_3 = 0.5(B_1 + B_2). \tag{74}$$

With

$$B_2 = B_1(1 + \delta), \tag{75}$$

it follows that

$$B_3 = B_1(1 + 0.5\,\delta). \tag{76}$$

Normalizing the filter gain by

$$\int h_n^2(x)\,dx = 4\pi A_n^2 B_n^2 \int e^{-2(\pi B_n x)^2}\,dx$$

$$= 2\sqrt{2\pi}\,A_n^2 B_n = \text{constant}, \tag{77}$$

the amplitudes A_n are related by

$$A_2 = A_1(1 + \delta)^{-0.5}, \tag{78}$$

$$A_3 = A_1(1 + 0.5\,\delta)^{-0.5}. \tag{79}$$

[21] S. O. Rice, "Statistical properties of a sine wave plus random oise," *Bell Sys. Tech. J.*, vol. 27, pp. 109–157; January, 1948. ee Appendix II.

Defining the noise components X_1, Y_1 and X_2, Y_2 with respect to the filter center frequencies f_n,

$$f_n - f_0 = 0 \tag{80}$$

and

$$\mu_1 = \mu_3 = 0, \tag{81}$$

$$\rho_2 = \pi^2 B_3^2(1 + 0.25\ \delta^2), \tag{82}$$

$$\rho_4 = 3\pi^4 B_3^4(1 + 1.5\ \delta^2), \tag{83}$$

where terms with powers higher than δ^2 have been neglected. Letting

$$\Omega = (f_1 - f_2)/B_3, \tag{84}$$

(48) gives

$$(\rho_a^2 + \mu_a^2)/\sigma^4 = 1 - 0.5\ \delta^2 - \Omega^2. \tag{85}$$

Computing the equivalent delay time τ from (34) or (35)

$$\tau^2\pi^2 B_3^2 = 0.5\ \delta^2 + \Omega^2, \tag{86}$$

where powers of δ and Ω above the second have been neglected.

B. Rectangular Filter

A rectangular band-pass filter of bandwidth B_n(cps) and of amplitude response A_n has an impulse response envelope

$$h_n(t) = 2A_n \sin \pi B_n t/(\pi t). \tag{87}$$

Computing the moments of the individual channels in Fig. 3(a) by (7)–(10), one has

$$\mu_{1a(n)} = 2\pi(f_n - f_0) + \cdots \tag{88}$$

$$\rho_{2a(n)} = \pi^2[B_n^2/3 + 4(f_n - f_0)^2] + \cdots \tag{89}$$

$$\mu_{3a(n)} = \pi^3(f_n - f_0)[2B_n^2 + 8(f_n - f_0)^2] + \cdots \tag{90}$$

$$\rho_{4a(n)} = \pi^4[0.2B_n^4 + 8(f_n - f_0)^2 B_n^2] + \cdots \tag{91}$$

Defining the reference bandwidth B_3 as in (74) and relating the bandwidths B_n by (75) and (76), normalizing the filter gain as in (77) gives

$$4A_n^2 B_n = \text{constant}. \tag{92}$$

The amplitudes A_n are related as in (78) and (79). Applying (80), the moment averages become

$$\mu_1 = \mu_3 = 0, \tag{93}$$

$$\rho_2 = (\pi^2 B_3^2/3)(1 + 0.25\ \delta^2), \tag{94}$$

$$\rho_4 = 0.2\pi^4 B_3^2(1 + 1.5\ \delta^2). \tag{95}$$

Applying (84), one gets from (48)

$$(\rho_a^2 + \mu_a^2)/\sigma^4 = \begin{cases} 1 - |\ \delta\ | + \frac{3}{8}\ \delta^2 \\ \qquad \text{for} \quad |\ f_1 - f_2\ |/B_1 < 0.5\ \delta \\ 1 - |\ 2\Omega\ | + \Omega^2 + \frac{1}{4}\ \delta^2 \\ \qquad \text{for} \quad |\ f_1 - f_2\ |/B_1 > 0.5\ \delta. \end{cases} \tag{96}$$

Computing the equivalent delay time τ from (34),

$$\pi^2 B^2 \tau^2 = \begin{cases} 3\left(|\ \delta\ | + \dfrac{\delta^2}{40}\right) \approx 3\ |\ \delta\ | \\ \qquad \text{for} \quad |\ f_1 - f_2\ |/B_1 < 0.5\ \delta \\ 6\left(|\ \Omega\ | + 0.3\Omega^2 - 0.125\ \delta^2\right) \approx 6\ |\ \Omega\ | \\ \qquad \text{for} \quad |\ f_1 - f_2\ |/B_1 > 0.5\ \delta. \end{cases} \tag{97}$$

C. A Triple-Tuned RLC Filter

A triple-tuned RLC filter is the simplest physically realizable filter for which the probability density $w(\phi, \ddot{\phi})$ may be defined.[9] A triple-tuned RLC filter of half-power bandwidth of B_n cps has an impulse response envelope

$$h_n(t) = A_n 2^{-0.5}(2\pi a_n)^3 t^2 e^{-2\pi a_n t}, \tag{98}$$

where

$$a_n = 0.98B_n. \tag{99}$$

Ignoring the two per cent difference in (99), B_n may be substituted for a_n in (98). Computing the moments of the individual channels in Fig. 3(a) by (55), (57), (59) and (60) one gets

$$\mu_{1a(n)} = 2\pi(f_n - f_0) + \cdots \tag{100}$$

$$\rho_{2a(n)} = \pi^2[4B_n^2/3 + 4(f_n - f_0)^2] + \cdots \tag{101}$$

$$\mu_{3a(n)} = 8\pi^3(f_n - f_0)[B_n^2 + (f_n - f_0)^2] + \cdots \tag{102}$$

$$\rho_{4a(n)} = 16\pi^4[B_n^4 + 2(f_n - f_0)^2 B_n^2] + \cdots \tag{103}$$

The bandwidths B_1, B_2 and B_3 are related as in (74) to (76). Normalizing the filter gain as in (77) gives

$$0.75\pi B_n A_n^2 = \text{constant}. \tag{104}$$

The amplitudes A_n are related as in (78) and (79). Applying (80), the moment averages become

$$\mu_1 = \mu_3 = 0, \tag{105}$$

$$\rho_2 = (4/3)\pi^2 B_3^2(1 + 0.25\ \delta^2), \tag{106}$$

$$\rho_4 = 16\pi^4 B_3^2(1 + 1.5\ \delta^2). \tag{107}$$

Eq. (48) gives

$$(\rho_a^2 + \mu_a^2)/\sigma^4 = 1 - 1.25(\delta^2 + \Omega^2). \tag{108}$$

Computing the equivalent time delay from (34) or (35),

$$\tau^2\pi^2 B^2 = \tfrac{15}{16}(\delta^2 + \Omega^2). \tag{109}$$

Appendix III
Accuracy Estimates

Two limiting difference frequency distributions may be obtained by selecting the moments μ_1, μ_3, ρ_2 and ρ_4 of the filters 1 or 2 of Fig. 3(a) as the moments of filter 3 in Fig. 3(b). The moments of the individual filters provide a moment-match between the two Gaussian distributions characterizing Figs. 3(a) and 3(b) that is inferior (in the least-squares sense) to the average moments of (26)–(29). Eqs. (26)–(29) may be expected to give a better accuracy

erence frequency distribution than the corresponding ments of filters 1 or 2. Although no proof of this will given, a heuristic argument may partially support the ove statement.

This explanation will be given for filters that differ in ndwidth, but whose relative skewness is constant. As g as the two filters have a power response symmetrical out the center of the filter pass band, constant relative ewness implies that the ratio $\Omega_n = \Delta_n/2\pi B_n = \ldots - \omega_0)/2\pi B_n$ remains constant. (ω_n = center frequency the filter, noise components X_n and Y_n are defined with pect to ω_0, B_n = filter bandwidth.) The moments μ_1 d μ_3, ρ_2 and ρ_4 are then proportional to powers of ,[22] and (5) may be rewritten as

$$w(\dot{\phi}, \ddot{\phi}) = B_n^{-3} f\left(\frac{\dot{\phi}}{B_n}, \frac{\ddot{\phi}}{B_n^2}\right). \tag{110}$$

tegrating (110) gives

$$w(\ddot{\phi}) = B_n^{-2} g(\ddot{\phi}/B_n^2). \tag{111}$$

tegrating (111) gives

$$P(|\ddot{\phi}| < \ddot{\phi}_0) = \int_{-\ddot{\phi}_0}^{\ddot{\phi}_0} B_n^{-2} g\left(\frac{\ddot{\phi}}{B_n^2}\right) d\ddot{\phi} = h\left(\frac{\ddot{\phi}_0}{B_n^2}\right). \tag{112}$$

pplying (21), one has

$$P(|\Delta\dot{\phi}| < \Delta\dot{\phi}_0, \tau) = h\left(\frac{\Delta\dot{\phi}_0}{\tau B_n^2}\right). \tag{113}$$

With $\tau \sim B_n^{-1}$ (this follows from (34) if the moments e proportional to powers of B_n),

$$P(|\Delta\dot{\phi}| < \Delta\dot{\phi}_0, \tau) = h\left(\frac{\Delta\dot{\phi}_0}{kB_n}\right) = h_1\left(\frac{\Delta\dot{\phi}_0}{B_n}\right). \tag{114}$$

[22] This can be seen from the expressions for moments given in pendix II.

For $P(|\Delta\dot{\phi}| < \Delta\dot{\phi}_0, \tau)$ = constant, (115)

$$\frac{\Delta\dot{\phi}_0}{B_n} = \text{constant} \tag{116}$$

or

$$\Delta\dot{\phi}_0 \sim B_n. \tag{117}$$

The wider spread of difference frequencies is associated with the wider filter. The intermediate filter bandwidth $B_3 \approx 0.5\ (B_1 + B_2)$ in Fig. 3(b) has a distribution of difference frequencies that is intermediate between those of filters 1 and 2.

It must still be shown that the difference frequency distribution in Fig. 3(a) approximates the difference frequency in Fig. 3(b) if $B_3 \approx 0.5\ (B_1 + B_2)$. The selection of τ in (34) ensures the same amount of correlation between noise components X_1Y_1 and X_2Y_2 in Fig. 3(a) and in Fig. 3(b), regardless of the bandwidth of the filter No. 3. The difference frequency in the output of Fig. 3(a) is the difference of the phase derivatives $\dot{\phi}$ between two correlated pairs of variables X_1Y_1 and X_2Y_2, one of which exhibits a wider $\dot{\phi}$ spread than the other.

The difference frequency fluctuations in the output of Fig. 3(b) are measured between similarly correlated pairs of variables X_1Y_1 and X_2Y_2, the frequency spread of which depends on the bandwidth B_3. The narrower filter of Fig. 3(a) when used as filter No. 3 would provide too small frequency fluctuations and hence differences. The wider filter would provide too large frequency differences; the filter bandwidth

$$B_3 = 0.5(B_1 + B_2) \tag{118}$$

may be expected to approximate the difference frequencies of Fig. 3(a).

Complementary Series*

MARCEL J. E. GOLAY†, FELLOW, IRE

Summary—A set of complementary series is defined as a pair of equally long, finite sequences of two kinds of elements which have the property that the number of pairs of like elements with any one given separation in one series is equal to the number of pairs of unlike elements with the same given separation in the other series.

(For instance the two series, 1001010001 and 1000000110 have, respectively, three pairs of like and three pairs of unlike adjacent elements, four pairs of like and four pairs of unlike alternate elements, and so forth for all possible separations.)

These series, which were originally conceived in connection with the optical problem of multislit spectrometry, also have possible applications in communication engineering, for when the two kinds of elements of these series are taken to be $+1$ and -1, it follows immediately from their definition that the sum of their two respective autocorrelation series is zero everywhere, except for the center term. Several propositions relative to these series, to their permissible number of elements, and to their synthesis are demonstrated.

Introduction and Definition

A SET OF complementary series is defined as a pair of equally long, finite sequences of two kinds of elements which have the property that the number of pairs of like elements with any given separation in one series is equal to the number of pairs of unlike elements with the same separation in the other series.

(For instance, the two series $A = 00010010$ and $B = 00011101$ are complementary.)

Series having the complementary property defined above were conceived originally in connection with the optical problem of infrared multislit spectrometry discussed in former publications.[1,2] In this application, which will be recalled briefly, almost the entire available field of the spectrometer is utilized, as illustrated by Fig. 1, instead of limiting its utilization to single entrance and exit slits, as is done in the ordinary spectrometer. This field is divided into four entrance portions, a, a', b and b', and the four exit portions into which the entrance portions are exactly imaged (with inversion) by radiation of the "proper" wavelength, *i.e.*, by the radiation to be measured (Fig. 1).

The a portion consists of a series of open or "closed" slits, from 40 to 200 in numbers, each slit being open or closed in accordance as to whether the corresponding element of the first of a pair of complementary series, A, and B, is a "one" or a "zero." In the a' portion, the open slits of a are closed slits, and vice versa. The b portion is made up likewise of open or closed slits in accordance with the B series, and in the b' portion the open slits of b are closed, and vice versa.

When the left or entrance half of the square field illustrated by Fig. 1 is uniformly illuminated by polychromatic radiation, which is imaged as a spread spectrum onto the right or exit half of the field, radiation of the proper wavelength images all the open slits of a and a' onto the corresponding open slits of a and a' at the exit half, and the radiation so passed contributes to the output D_1 of a detector designed to measure the output of the a and a' exit portions. Likewise, all of the radiation of that same wavelength passed by the b and b' portions of the entrance field will be exactly blocked by the exit portions b' and b, respectively, while the remaining radiation exiting through the b' and b portion will be measured by a detector giving an output D_2.

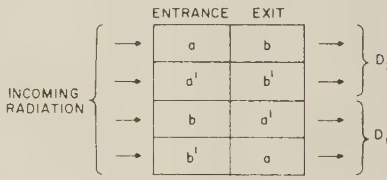


Fig. 1—Field utilization in multislit spectrometer.

For radiation of a different wavelength causing the image of the left half of the field to be shifted j slitwidths on the exit half, the number of open pairs of slits permitting the $a - a$ and the $a' - a'$ passage will be equal to the number of pairs of j-separated like elements in the A series. On the other hand, the number of open pairs of slits permitting the $b - b'$ and the $b' - b$ passage will be equal to the number of pairs of j-separated unlike elements in the B series. Since this number equals the number of j-separated like elements in the A series, it follows that the difference $D_1 - D_2$ between the measures of the two radiation bundles passing the exit half of the field will be unaffected by any radiation except that of the proper wavelength, and will constitute a measure of much more radiation of that proper wavelength than if single entrance and exit slits had been utilized.

The basic property of complementary series may be expressed also in autocorrelative terms. Let the various a_i and b_i elements ($i = 1, 2, \cdots n$) of two n-long complementary series be either $+1$ or -1, and let their respective autocorrelative series be defined by

$$c_j = \sum_{i=1}^{i=n-j} a_i a_{i+j}$$

* Received by the PGIT, May 23, 1960; revised manuscript received September 20, 1960.

† Philco Corp., Philadelphia, Pa.

[1] M. J. E. Golay, "Multislit spectrometry," *J. Opt. Soc. Am.*, vol. 39, p. 437; 1949.

[2] M. J. E. Golay, "Static multislit spectrometry and its application to the panoramic display of infrared spectra," *J. Opt. Soc. Am.*, vol. 41, p. 468; 1951.

d

$$d_j = \sum_{i=1}^{i=n-j} b_i b_{i+j}.$$

e have

$$c_j + d_j = 0 \qquad j \neq 0$$

$$c_0 + d_0 = 2n.$$

his autocorrelative property of complementary series ay lead to applications in the field of communication in -called horizontal modulation systems, which permit veral communication channels to utilize simultaneously e same frequency bands. These modulation systems are quiring increasing importance.

Regardless of past or possible future applications, the riter has found these complementary series mathnatically appealing, first because of the deep seated mmetries which characterize them, even though no gn of order may be obvious at first glance, and second ecause of the challenge offered by the problem of synesizing them for $n = 26$, 34, etc.

Conventions and Terminology

Two pairs of elements will be termed like pairs when oth are pairs of like elements or when both are pairs of like elements; otherwise they will be termed unlike irs.

Wherever convenient, pairs of like or unlike elements ill be termed even and odd pairs, respectively, and a ad of four elements will be termed even or odd dependg upon whether this quad can be decomposed into two e pairs or two unlike pairs.

Pairs of elements which are distant an even number of ements will be termed even spaced pairs, and the others ill be termed odd spaced pairs.

Whenever each element of a set is replaced by an ement of the other kind, it will be said that these eleents and the set are altered, and this operation will be dicated by a prime.

In all that follows, the two kinds of elements will be the , 1} set. Thus the parity of a pair or of a quad will be mply the sum of its elements modulo 2, and we shall ve $a' \equiv a + 1 \pmod 2$.

neral Properties

1) The numbers of elements in two complementary ries are equal. If it were not so, the pair of extreme ments of the longest series would remain unmatched an unlike pair of elements with the same spacing in e other series.

2) Two complementary series are interchangeable. It ll be noted that this results from the symmetry of the finition with respect to the A and B series.

3) The order of the elements of either or both of a pair complementary series may be reversed. This results om the circumstance that the order of a pair of elements es not affect the parity of this pair.

4) One or both of a pair of complementary series may be altered without affecting their complementary property. This results from the circumstance that the parity of a pair is invariant under alteration of both elements of that pair.

5) Alternate elements in each of two complementary series may be altered, without affecting their complementary property. Such a transformation results in the change of both or neither elements of an even spaced pair, so that the parity of such pairs remains unaffected. Conversely, the parity of the odd spaced pairs is changed in both series, and this, by virtue of the remarks made in 2), does not affect the complementary property of the series.

It is concluded from the properties 2)–5) that a single pair of complementary series can be the basis for the construction of 2^6 pairs of complementary series (some of which may be identical) by either performing or not performing the following six operations:

a) Interchanging the series.
b) Reversing the first series.
c) Reversing the second series.
d) Altering the first series.
e) Altering the second series.
f) Altering the elements of even order of each series.

Examples

The six individual operations listed above, when performed one at a time on the complementary series given above, yield the six new pairs of complementary series:

0 0 0 1 1 1 0 1 and 0 0 0 1 0 0 1 0

0 1 0 0 1 0 0 0 and 0 0 0 1 1 1 0 1

0 0 0 1 0 0 1 0 and 1 0 1 1 1 0 0 0

1 1 1 0 1 1 0 1 and 0 0 0 1 1 1 0 1

0 0 0 1 0 0 1 0 and 1 1 1 0 0 0 1 0

0 1 0 0 0 1 1 1 and 0 1 0 0 1 0 0 0.

It will be noted that performing successively operations a)–e) on the original example will, in this particular case reproduce the original pair:

0 0 0 1 1 1 0 1 and 0 0 0 1 0 0 1 0

1 0 1 1 1 0 0 0 and 0 0 0 1 0 0 1 0

1 0 1 1 1 0 0 0 and 0 1 0 0 1 0 0 0

0 1 0 0 0 1 1 1 and 0 1 0 0 1 0 0 0

0 0 0 1 0 0 1 0 and 0 0 0 1 1 1 0 1.

It can be verified by inspection that no combination of the 6 operations listed above, with each operation performed at most once, and in the order listed, will reproduce the following complementary series:

1 0 0 1 0 1 0 0 0 1 and 1 0 0 0 0 0 0 1 1 0.

6) When the complementary property is written explicitly for the two pairs which are $n-1$ elements distant in the A and B series, we obtain

$$a_1 + a_n + b_1 + b_n \equiv 1 \pmod{2}. \tag{1}$$

When the complementary property is written for the four pairs which are $n-2$ elements distant, we obtain

$$a_1 + a_2 + a_{n-1} + a_n + b_1 + b_2 + b_{n-1} + b_n \equiv 0 \pmod{2}, \tag{2}$$

and by addition modulo 2 of (1) and (2)

$$a_2 + a_{n-1} + b_2 + b_{n-1} \equiv 1 \pmod{2}. \tag{3}$$

The process may be continued to show that, generally

$$a_r + a_{n-r+1} + b_r + b_{n-r+1} \equiv 1 \pmod{2}. \tag{4}$$

When $n = 2s + 1$ and $r = s + 1$, substitution in (4) yields

$$a_{s+1} + a_{s+1} + b_{s+1} + b_{s+1} \equiv 1 \pmod{2}$$

which is self-contradictory. Hence, it is concluded that the number of elements in complementary series must be even.

7) Let

$$u(x, y) = (x - y)^2, \qquad x = 0 \quad \text{or} \quad 1, \qquad y = 0 \quad \text{or} \quad 1,$$

the function of x and y thus defined is 0 or 1 depending upon the xy pair being even or odd.

We shall have, for complementary series

$$\sum_{s=1}^{s=v} u(a_s, a_{n-v+s}) + u(b_s, b_{n-v+s}) = v; \tag{5}$$

that is, the total number of odd pairs of elements which are $n - v$ elements distant in two complementary series is v. That is also, of course, the total number of even pairs of elements which are spaced likewise.

Now let

$$t(v) = \tfrac{1}{2} \sum_{s-1}^{s=v} (a_s + a_{n-v+s} + b_s + b_{n-v+s}) - \tfrac{1}{2}v$$

$$= \tfrac{1}{2} \sum_{s=1}^{s=v} (a_s + a_{n-s+1} + b_s + b_{n-s+1}) - \tfrac{1}{2}v. \tag{6}$$

The terms under each sum appear the same number of times as in the LHS of (5); and since there are v 1's among them which must be associated each with a 0 to make v odd pairs, one half the excess of the 1's included under the $\sum$ sign over v represents the number of pairs of 1's which are $n - v$ elements distant. The last member of (6) may be utilized therefore to determine how many pairs of 1's there should be with any given spacing, among two complementary series, and this reduces to approximately one quarter the number of pairs which must be examined in order to verify the complementary property of two series.

Example

The two complementary series: 1000110110000010 and 0100000101001110 are written folded over, starting at the bottom, going up and then going down again, as follows:

1 1 1 0	2 1
0 0 0 1	3 1
1 0 0 0	3 1
1 0 0 0	3 1
0 0 0 1	3 1
0 0 0 1	3 1
0 1 1 1	3 0
1 0 0 0	4.

The numbers written at the right are the $t(v)$'s, and they are obtained in the reverse order by starting at 0 and adding 1 every time a row containing three 1's has been passed, going up and then going back down. It is immediately verified that there are four pairs of adjacent 1's, three pairs of alternate 1's, etc., up to one pair of 1's which are 14 elements distant.

8) Let p and q designate the numbers of 1's in two complementary series. Since the total number of even pairs in one must equal the total number of odd pairs in the other, we have

$$\tfrac{1}{2}p(p-1) + \tfrac{1}{2}(n-p)(n-p-1) = q(n-q)$$

whence

$$n = (n - p - q)^2 + (p - q)^2; \tag{7}$$

i.e., the number of elements in complementary series must be expressible as a sum of at most two squares. Since this number must also be even, the allowable numbers up to 50 are

2, 4, 8, 10, 16, 18, 20, 26, 32, 34, 36, 40, and 50.

General Synthesis

9) Consider the two series

$$S_1 = AB = a_1 \cdots a_n b_1 \cdots b_n$$

$$S_2 = AB' = a_1 \cdots a_n b_1' \cdots b_n' \tag{8}$$

formed by appending the series A and B, and the series A and B'.

It will be noted that the number of pairs of like a elements with any given spacing in the first is equal to the number of pairs of unlike b' elements with the same given spacing in the second, and that the number of pairs of like b elements with any given spacing in the first is equal to the number of pairs of unlike a elements with the same given spacing in the second. Furthermore, to any ab pair of elements in the first corresponds the homologous unlike ab' pair of elements written immediately below in the second. This accounts for all pairs of elements in both S_1

nd S_2 series, and it is concluded therefore that these ries are also complementary.

10) It can be shown with a similar reasoning that the terleaved series

$$T_1 = a_1 b_1 a_2 b_2 \cdots a_n b_n$$

nd

$$T_2 = a_1 b_1' a_2 b_2' \cdots a_n b_n' \tag{9}$$

re also complementary.

11) Let $C = c_1 \cdots c_m$ and $D = d_1 \cdots d_m$ designate nother pair of complementary series, and consider the wo series

$$U_1 = A^{c_1} A^{c_2} \cdots A^{c_m} B^{d_1} B^{d_2} \cdots B^{d_m}$$

nd

$$U_2 = A^{d_m} \cdots A^{d_2} A^{d_1} B^{c_m \prime} \cdots B^{c_2 \prime} B^{c_1 \prime} \tag{10}$$

here the parity of the exponent determines whether the or B subseries is altered (odd exponent) or not (even xponent).

It will be noted first that the number of pairs of like lements with any given spacing within any A or B ubseries of the U_1 series is matched by the number of airs of unlike elements with the same spacing in the espective B or A subseries of the U_2 series.

It will be noted next that any pair of elements, with any iven spacing, one taken in the A^{c_i} subseries and the other the B^{d_i} subseries, is matched by a homologous unlike air of elements with the same spacing one taken in the ^{d_i} subseries and the other in the B^{c_i} subseries.

It will be noted next that all pairs of elements, one taken subseries A^{c_i} and the other in subseries $A^{c_{i+k}}$ of he U_1 series, can be matched with exactly as many unlike airs of elements as are taken, one in subseries A^{d_i} and he other in subseries $A^{d_{i+k}}$ of the U_2 series. This llows immediately from the complementary character the C and D series.

And last, the preceding observation will be made also bout the pairs of elements selected from different B ries. This exhausts the ensemble of the pairs of elements the U_1 and U_2 series, and it is concluded that the U_1 nd U_2 series are complementary.

12) It can be shown with a similar reasoning that the terleaved series

$$V_1 = A^{c_1} B^{d_1} A^{c_2} B^{d_2} \cdots A^{c_m} B^{d_m}$$

nd

$$V_2 = A^{d_m} B^{c_m \prime} \cdots A^{d_2} B^{c_2 \prime} A^{d_1} B^{c_1 \prime} \tag{11}$$

here the symbols of the RHS have the same connotations in (10), are also complementary series.

It is concluded from 10), and also from 11), that, given vo pairs of complementary series of n and m elements, spectively, a pair of complementary series with $2nm$ ements can be synthesized therefrom.

Complementary Series in Which n is a Power of 2

13) Let the generalized boolean sum $b(a_i, a_j)$ be constrained by

$$b(0, 0) + b(0, 1) + b(1, 0) + b(1, 1) \equiv 1 \pmod{2}. \tag{12}$$

Let $x_1 x_2 \cdots x_e$ designate the number x written in the binary system. Let

$$e_1(x) \equiv b(x_{\alpha_1}, x_{\alpha_2}) + b(x_{\alpha_2}, x_{\alpha_3}) + \cdots + b(x_{\alpha_{e-1}}, x_{\alpha_e}) \pmod{2}$$

and

$$e_2(x) \equiv e_1(x) + x_{\alpha_1} \pmod{2} \tag{13}$$

where the subscripts $\alpha_1, \alpha_2, \cdots \alpha_e$ designate any permutation of the numbers $1, 2, \cdots e$. It will be shown that the two series

$$E_1 = e_1(0), e_1(1), \cdots e_1(2^{e-1})$$

and

$$E_2 = e_2(0), e_2(1), \cdots e_2(2^{e-1})$$

are complementary.

Let x^1 and y^1 designate two distinct numbers with the respective binits $x_1^1, x_2^1, \cdots x_e^1$ and $y_1^1, y_2^1, \cdots y_e^1$, and associate with x^1 and y^1 the numbers x^2 and y^2 formed by changing the binit preceding the first binit, in the order defined by α_1, α_2, etc., which is different in x^1 and y^1. This association is clearly reciprocal and univocal, and we shall have always

$$y^1 - x^1 = y^2 - x^2 \tag{14}$$

since the change made in x^1 and y^1 either adds or subtracts the same power of 2 to or from x^1 and y^1.

Consider now the pair of elements defined by x^1 and y^1 in the first series, and associate it with the pair of elements defined by x^2 and y^2 in the second series. If

$$x_{\alpha_1}^1 \neq y_{\alpha_1}^1, \tag{15}$$

there is no preceding binit which can be changed and we shall have

$$x^2 = x^1$$
$$y^2 = y^1. \tag{16}$$

From (13) and (15) we derive

$$e_1(x^1) + e_1(y^1) + e_2(x^2) + e_2(y^2) \equiv x_{\alpha_1}^1 + y_{\alpha_1}^1 \equiv 1 \pmod{2} \tag{17}$$

which indicates that the two pairs associated with each other are unlike for the case defined by (15).

If on the other hand, the first binits which are different in x^1 and y^1 are the $x_{\alpha_i}^1$ and $y_{\alpha_i}^1$ binits

$$x_{\alpha_i}^1 \neq y_{\alpha_i}^1, \qquad i > 1, \tag{18}$$

we shall have

$$x^1_{\alpha_i} = x^2_{\alpha_i}, \quad y^1_{\alpha_i} = y^2_{\alpha_i} \tag{19}$$

$$x^1_{\alpha_{i-1}} = y^1_{\alpha_{i-1}} \neq x^2_{\alpha_{i-1}} = y^2_{\alpha_{i-1}} \tag{20}$$

$$x^1_{\alpha_{i-2}} = y^1_{\alpha_{i-2}} = x^2_{\alpha_{i-2}} = y^2_{\alpha_{i-2}} \quad \text{when} \quad i > 2 \tag{21}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$x^1_{\alpha_1} = y^1_{\alpha_1} = x^2_{\alpha_1} = y^2_{\alpha_1}. \tag{22}$$

The calculation of the parity of the $x^1y^1x^2y^2$ quad needs involve only those elements which are formally different in e_1 and e_2:

$$\begin{aligned} e_1(x^1) &+ e_1(y^1) + e_2(x^2) + e_2(y^2) \\ &= b(x^1_{\alpha_{i-2}}, x^1_{\alpha_{i-1}}) + b(x^1_{\alpha_{i-1}}, x^1_{\alpha_i}) + b(y^1_{\alpha_{i-2}}, y^1_{\alpha_{i-1}}) \\ &\quad + b(y^1_{\alpha_{i-1}}, y^1_{\alpha_i}) + b(x^2_{\alpha_{i-2}}, x^2_{\alpha_{i-1}}) + b(x^2_{\alpha_{i-1}}, x^2_{\alpha_i}) \\ &\quad + b(y^2_{\alpha_{i-2}}, y^2_{\alpha_{i-1}}) + b(y^2_{\alpha_{i-1}}, y^2_{\alpha_i}) + x^2_{\alpha_1} + y^2_{\alpha_1}. \end{aligned} \tag{23}$$

The last two terms in the RHS of (23) cancel by virtue of (22). Furthermore, either the terms involving α_{i-2} do not exist, when $i = 2$, or the first and third terms of the RHS cancel by virtue of (20) and (21), and so do the fifth and seventh terms. The remaining terms may be written, by virtue of (12), (18), (19) and (20).

$$\begin{aligned} b(x^1_{\alpha_{i-1}}, x^1_{\alpha_i}) &+ b(x^1_{\alpha_{i-1}}, x^1_{\alpha_i} + 1) + b(x^1_{\alpha_i} + 1, x^1_{\alpha_i}) \\ &+ b(x^1_{\alpha_{i-1}} + 1, x^1_{\alpha_i} + 1) \equiv 1 \pmod{2}. \end{aligned} \tag{24}$$

Thus it has been shown that every pair of elements of the first series can be associated with an unlike pair of elements in the second series which, by virtue of (14), have the same spacing; and that the association is univocal and reciprocal. This concludes the proof that the series, the individual elements of which are defined by (13), are complementary.

14) Since the order of the α's in (13) may be reversed without affecting the argument which follows, the two series

$$e_1x \equiv b(x_{\alpha_1}, x_{\alpha_2}) + b(x_{\alpha_2}, x_{\alpha_3}) + \cdots + b(x_{\alpha_{e}-1}, x_{\alpha_e}) \pmod{2}$$

and

$$e_2^*(x) = e_1(x) + x_{\alpha_e} \tag{25}$$

are also complementary.

15) The constraint (12) remains valid when $b(x_{\alpha i}, x_{\alpha j})$ is replaced by $b(x_{\alpha i}, x_{\alpha j}) + x_{\alpha i}$ or $b(x_{\alpha i}, x_{\alpha j}) + x_{\alpha j}$, or $b(x_{\alpha i}, x_{\alpha j}) + x_{\alpha j} + x_{\alpha j}$.

Therefore, $e_1(x)$ and $e_2(x)$, and also $e_1(x)$ and $e_2^*(x)$, remain pairs of complementary series when the sum

$$\sum a_i x_i, \qquad a_i = 0 \quad \text{or} \quad 1 \tag{26}$$

is added modulo 2 to both $e_1(x)$ and $e_2(x)$, or both $e_1(x)$ and $e_2^*(x)$.

The series formed by calculating the expression (26) modulo 2 for any sequence of numbers $0, 1 \cdots 2^e - 1$, will be termed Walsh series by analogy with the Walsh functions which are obtained when 0 is replaced by -1 in these series, and the result obtained above may be stated as follows:

The complementary series defined by (13), and also those defined by (26) retain their complementary property when they are added, element by element and modulo 2, to a Walsh series having the same number of elements.

16) From the definition (13) may be derived the following construction, when passing from series with 2^e elements to series with 2^{e+1} elements:

When two complementary series have been formed in accordance with (13), the first series with 2^{e+1} elements formed by interlacing the successive subseries of 2^f elements into which the two complementary series with 2^e elements may be divided, and the second series with 2^{e+1} elements formed by altering alternate subseries of 2^f elements in the first series with 2^{e+1} elements, are complementary series.

Complementary Series in Which n is not a Power of 2.

$n = 10$. Two basic pairs of complementary series exist for this case, from which all others may be derived by performing one or more of the six operations listed in 5). These two pairs of series are:

1 0 0 1 0 1 0 0 0 1	and	0 1 0 1 0 0 0 0 1 1,
1 0 0 0 0 0 0 1 1 0		0 0 0 0 1 0 0 1 1 0
9 6 3 10 7 4 1 8 5 2		7 10 3 6 9 2 5 8 1 4.

A relationship exists between these two pairs, which has been indicated by the numbers under them. Starting with the seventh pair of symbols of the left pair of series, and selecting every third pair of symbols thereafter, always recycling when the end of the series is reached, reproduces the successive pairs of symbols of the right pair of series. The left pair of series may be obtained similarly from the right pair, as indicated by the numbers below the right pair of series.

From these two basic pairs, complementary series with $10.2^a .20^b$ elements may be derived, utilizing the synthesis methods described earlier.

$n = 18$. It has been verified by trial that complementary series do not exist for $n = 18$.

$n = 26$. An extensive, yet not exhaustive "longhand" search has not disclosed any complementary series for this case.

Conclusion

When n is a power of 2, general methods have been etermined, which lead to the formation of a large multi-icity of complementary series, albeit it has not been iown that these methods constitute the most general ethods.

When n is not a power of 2, complementary series have een discovered for the basic case $n = 10$ only. They have een verified not to exist for $n = 18$.

The case $n = 26$ appears too laborious for an exhaustive nghand search. It is one purpose of this discussion to cpress the hope that someone interested in number or roup theory and having access to an electronic computer ill have the curiosity to program a computer to make an chaustive search for this case, and that if solutions are und, they may permit some generalizations, and the establishment of productive connections between these series and number or group theory.

The next case for a nonpower of 2 is that of $n = 34$. Utilizing the complementary properties expressed by (4), and also the properties embodied in the 6 operations listed earlier, we may reduce the 68 choices of elements to $\frac{3}{2}\,34 - 6 = 45$ binary choices.

Thus 2^{45} combinations should be investigated, or a somewhat reduced number, if more elaborate properties, not discussed above, are utilized.[3] Even when so reduced, the numbers involved still appear formidable for an electronic computer.

[3] For instance, if complementary series are sought for $n = 34$, and if we set $p = 13$ and $q = 16$ in (7), the series with 13 1's can be shown to consist of one series of 17 elements with 6 1's interleaved with another series of 17 elements with 7 1's; likewise, the series with 16 1's can be shown to be the interleaving of 2 series with 6 and 10 1's, respectively.

Signal Detection by Adaptive Filters*

E. M. GLASER†, MEMBER, IRE

Summary—**Communication engineers are now giving increased tention to detection systems which are able to adjust their own ructure so as to be optimum for the particular detection problem the moment. This paper describes a system which is capable of lapting and optimizing its response to the class of pulse signals hose individual pulses are less than *T* seconds in duration.**

The analysis and synthesis of the adaptive system is facilitated the use of an orthogonal function decomposition of the received gnal. The use of the orthogonal decomposition permits synthesis optimum linear filters by various circuit techniques, several of hich have been reported elsewhere. The structure of the system ilizing such a decomposition is described in detail.

Since the operation of the adaptive filter is based upon signal etection and estimation in noise backgrounds, considerable at-ntion is devoted to the relationship between optimum signal etection and estimation. The methods of statistical decision theory e used.

A program to test the validity of the approximations and assess e over-all system performance was carried out by simulation of e system on both analog and digital computers. The results of ese experimental runs are described.

* Received by the PGIT, June 2, 1960; revised manuscript ceived, September 26, 1960. This research was supported by the SAF through the Wright Air Dev. Div. of the Air Res. and Dev. ommand.

† Radiation Lab., The Johns Hopkins University, Baltimore Md.

Introduction

MOST of the work in communication engineering concerning weak signal detection has been devoted to the study of the synthesis and performance of optimum, time invariant filters and detectors, that is, to detection systems whose structure is fixed in a configuration which is optimum for the particular signal or class of signals that is to be received [1]. In order to design a detection system of this class and expect it to yield any degree of satisfactory performance, the signal waveforms must be known (together with their statistics, if the waveforms vary), as must the statistical properties of the assumed stationary interfering noise. The designer's restriction to an invariant system is one of the reasons for this being so. This fact is now well recognized, and increasing attention is being given to detection systems which are able to adapt their structure so as to be optimum for the particular detection problem of the moment. Examples of such systems are discussed in detail in [2] and [3]. These systems are restricted to the reception of a particular signal whose structure is known, but not its time of arrival, or epoch. The type of interference is known to be additive Gaussian noise of constant average power and slow multipath signal fading.

This paper describes a type of adaptive detection system suitable for the reception of a pulse signal whose waveform is fixed, but unknown at the receiver. The system is one which functions initially as an incoherent detector and, as it receives more and more pulses from a particular source, modifies its structure and optimizes its detection performance with respect to this signal.

Optimum Detection of Signals Via Decision Theory

A substantial amount of the work performed in this report relies upon the ideas and theorems of matched filter theory[1] and decision theory [4], [8]. Decision theory treats in a general way the problem relating to the detection of signals in noise and the estimation of their structure. The approach is based upon the fact that in testing hypotheses all decisions involve doubt and uncertainty and have associated with them various costs and risks. These may be measured in any way appropriate to the problem at hand. The cost is taken to be a function both of the true hypothesis and the hypothesis as the observer decides it to be. The conditional risk is the average value of the cost over all possible decisions the observer can make given a particular hypothesis. The average risk is the average of the conditional risk over all possible hypotheses. Decision theory assumes that the observer wishes to behave in the way which will minimize his conditional or average risk. On this assumption, it shows the observer how to choose a decision rule for processing the received data. This decision rule will yield decisions which minimize risk for the particular physical situation and the costs involved.

It is known from decision theory that a wide class (Bayes') of tests for the optimum detection of signals in noise is based upon the use of the likelihood ratio Λ, given by

$$\Lambda = \frac{Y(\mathbf{S})F(\mathbf{V} \mid \mathbf{S})}{Y(0)F(\mathbf{V} \mid 0)}. \tag{1}$$

The numerator of this ratio is the joint probability of data $\mathbf{V}$ and signal $\mathbf{S}$; the denominator is the joint probability of data $\mathbf{V}$ and the zero signal 0.

Y is the *a priori* distribution of received signal $\mathbf{S}$ in a signal vector space Ω. $\mathbf{V}$ is the received data vector in vector space Γ, and F is the conditional probability distribution of $\mathbf{V}$ given $\mathbf{S}$. $\mathbf{V}$ is taken to be the sum of signal and noise $\mathbf{N}$. A signal is said to be present whenever Λ exceeds some preset threshold level. The value of this threshold is dependent upon the test chosen and the various costs involved. Log Λ, a monotonically increasing function of Λ, is more convenient mathematically and physically to work with. The detection threshold for log Λ will be denoted by W.

[1] See the Matched Filter issue of IRE Trans. on Information Theory, vol. IT-6, June, 1960. In particular, the article by G. Turin, "An Introduction to Matched Filter Theory," is useful.

It is convenient for our purposes to represent the received signal $S(t)$ by an expansion of the form

$$S(t) = \sum_{j=1}^{\infty} s_j\phi_j(t), \tag{2}$$

and the vector $\mathbf{S} = (s_1, s_2, \cdots, s_j \cdots)$. The ϕ_j are orthonormal on an interval which completely spans the interval $(0, T)$ in which all the $S(t)$ of interest are assumed to exist. We note here that we are interested in detecting single pulses received from a source which emits pulses of a constant shape,[2] but not necessarily with constant repetition rate. The received pulses need not have the same shape as the transmitted pulses. The set of $\phi_j(t)$ which is of interest here is that which is a solution of the integral equation

$$\int_0^T K_N(t, u)\phi_j(u)\, du = \sigma_j^2\phi_j(t) \qquad 0 \leq (t, u) \leq T.$$

$$j = 1, 2, 3, \cdots. \tag{3}$$

The kernel $K_N(t, u)$ is the autocorrelation function of the noise, $N(t)$. $N(t)$ can be expanded in the form $\sum_{j=1}^{\infty} n_j\phi_j(t)$. With this set of functions, $E(n_jn_k) = 0$ and $E(n_j^2) = \sigma_j^2$, so that, for Gaussian statistics, the noises $n_j\phi_j(t)$ and $n_k\phi_k(t)$ are statistically independent.

There are two detection situations which are of particular interest:

1) Incoherent detection: The *a priori* probability of a nonzero signal is p, and the probability distribution of such a nonzero signal is uniform over Ω.

2) Coherent detection: The signal waveform is known to be exactly $\mathbf{S}_0$. The *a priori* probability distribution of signals is then a Dirac delta function: $Y(S) = \delta(\mathbf{S} - \mathbf{S}_0)$.

Incoherent Detection

In this situation it can be shown that when the noise is Gaussian, the log of the likelihood ratio $\Lambda_{(i)}$ is given by

$$\log \Lambda_{(i)} = \log \Lambda_0^{(i)} + \tfrac{1}{2} \sum_{j=1}^{\infty} \frac{v_j^2}{\sigma_j^2}, \tag{4}$$

with the constant term

$$\log \Lambda_0^{(i)} = \log \mu - \sum_{j=1}^{\infty} \log \sigma_j \sqrt{2\pi} \tag{5}$$

and

$$\mu = p/q = p/1 - p. \tag{6}$$

When $\sigma_j = \sigma$, (4) can be reduced to

$$\log \Lambda_{(i)} = \log \Lambda_0^{(i)} + \frac{1}{2\sigma^2} \int_0^T V^2(t)\, dt, \tag{7}$$

which is the well-known result that energy detection is the optimum form of detection when signal structure is unknown.

[2] Pulses $f_1(t)$ and $f_2(t)$ are of constant shape if $f_2(t) = kf_1(t \pm \tau)$.

The Bayes' optimum quadratic cost function estimator or $\mathbf{S}$ is given by (8):[3]

$$\mathbf{S}^* = \frac{\int_\Omega \mathbf{S} Y(\mathbf{S}) F(\mathbf{V} \mid \mathbf{S})\, d\mathbf{S}}{\int_\Omega Y(\mathbf{S}) F(\mathbf{V} \mid \mathbf{S})\, d\mathbf{S}}. \tag{8}$$

When Y is uniform, this can be reduced [5] to yield the optimum estimators for the coefficients of $\mathbf{S}$:

$$s_j^* = v_j = \int_0^T V(t)\phi_j(t)\, dt. \tag{9}$$

Coherent Detection

When the noise is Gaussian and signal $\mathbf{S}_0$ is received, the log of the likelihood ratio $\Lambda_{(c)}$ is given by

$$\log \Lambda_{(c)} = \log \Lambda_0^{(c)} + \sum_{j=1}^{\infty} \frac{s_{0j}v_j}{\sigma_j^2}, \tag{10}$$

with the constant term

$$\log \Lambda_0^{(c)} = \log \mu - \tfrac{1}{2} \sum_{j=1}^{\infty} \frac{s_{0j}^2}{\sigma_j^2}. \tag{11}$$

When $\sigma_j = \sigma$, (10) can be rewritten as

$$\log \Lambda_{(c)} = \log \Lambda_0^{(c)} + \frac{1}{\sigma^2} \int_0^T V(t)h(-t)\, dt \tag{12}$$

$$h(-t) = \sum_{j=1}^{\infty} s_{0j}\phi_j(t). \tag{13}$$

Thus, (12) is seen to yield the familiar matched filter.

Filter Systems for Incoherent and Coherent Detection

It should be noted that both $\Lambda_{(i)}$ and $\Lambda_{(c)}$ are functions of time, so that (4) and (10) are more completely written as

$$\log \Lambda_{(i)}(t) = \log \Lambda_0^{(i)} + \tfrac{1}{2} \sum_{j=1}^{\infty} \frac{v_j^2(t)}{\sigma_j^2} \tag{14}$$

and

$$\log \Lambda_{(c)}(t) = \log \Lambda_0^{(c)} + \sum_{j=1}^{\infty} \frac{s_{0j}}{\sigma_j^2} v_j(t). \tag{15}$$

The incoherent and coherent likelihood ratio computers or detectors can be synthesized by means of time-invariant linear filter systems. The linear filters have the impulses responses $\phi_j(-t)$. In any real system there are only a finite number of J of these, usually the first J of the set. It is also necessary to provide for epoch estimation, a means for determining when the likelihood ratio is at a maximum. This is easily done by differentiating log Λ and generating an epoch pulse whenever both d/dt log Λ passes through zero with negative slope and log Λ exceeds the detection threshold. Figs. 1 and 2 show block diagrams for the incoherent and coherent detection systems. The optimum component estimate outputs (obtained when log Λ is a maximum) are shown on these diagrams. Optimum estimates s_{oj}^* for the coherent system are also found by (9).

A Detection System for Gauss-Distributed Signals

When the signals are Gaussian distributed *a priori* in signal space, it is possible to specify another Bayes' optimum detector by use of (1). Suppose, in particular, that $Y(\mathbf{S})$ can be written

$$Y(\mathbf{S}) = y_1(s_1)y_2(s_2) \cdots y_j(s_j) \cdots, \tag{16}$$

y_j is Gaussian with mean $E(s_j)$ and variance $D^2(s_j)$. The signal coefficients are then statistically independent of one another *a priori*. This assumpton is a cautious one and somewhat unrealistic, for there is likely to be considerable interdependence among the parameters. Nevertheless, this type of distribution is a quite useful one to consider in connection with the process of adaptation to be discussed here. The likelihood ratio can be obtained via straightforward calculations and is given by[4]

$$\log \Lambda_{(g)}(t) = \log \Lambda_0^{(g)} + \tfrac{1}{2} \sum_{j=1}^{\infty} \frac{D^2(s_j)v_j^2(t)}{\sigma_j^2(\sigma_j^2 + D^2(s_j))} + \sum_{j=1}^{\infty} \frac{E(s_j)v_j(t)}{\sigma_j^2 + D^2(s_j)} \tag{17}$$

$$\log \Lambda_0^{(g)} = \log \mu + \sum_{j=1}^{\infty} \log \left[\frac{\sigma_j}{(\sigma_j^2 + D^2(s_j))^{1/2}} \right] - \tfrac{1}{2} \sum_{j=1}^{\infty} \frac{E^2(s_j)}{\sigma_j^2 + D^2(s_j)}. \tag{18}$$

It can be seen that this likelihood ratio computer is a combination of the incoherent and coherent detectors discussed above. The optimum parameter estimator can be obtained from (8) and is

$$s_j^* = \frac{D^2(s_j)v_j + E(s_j)\sigma_j^2}{\sigma_j^2 + D^2(s_j)}. \tag{19}$$

This estimator weights the data input in proportion to the *a priori* signal component variance and the *a priori* expected value in proportion to the noise variance. Fig. 3 is a block diagram of a detection and estimation system for signals having a Gauss *a priori* distribution.

The Adaptive Filter

It is now possible to describe an adaptive filter system for the detection and analysis of pulse signals. This system is designed for reception of pulse signals whose waveforms and amplitudes are fixed with time, but unknown *a priori* at the receiver. This amplitude restriction will be relaxed later in the paper. The system operates in such a way as to make the structural transition from

[3] See [4], Eq. (4.8).

[4] See [5], Eq. (95).

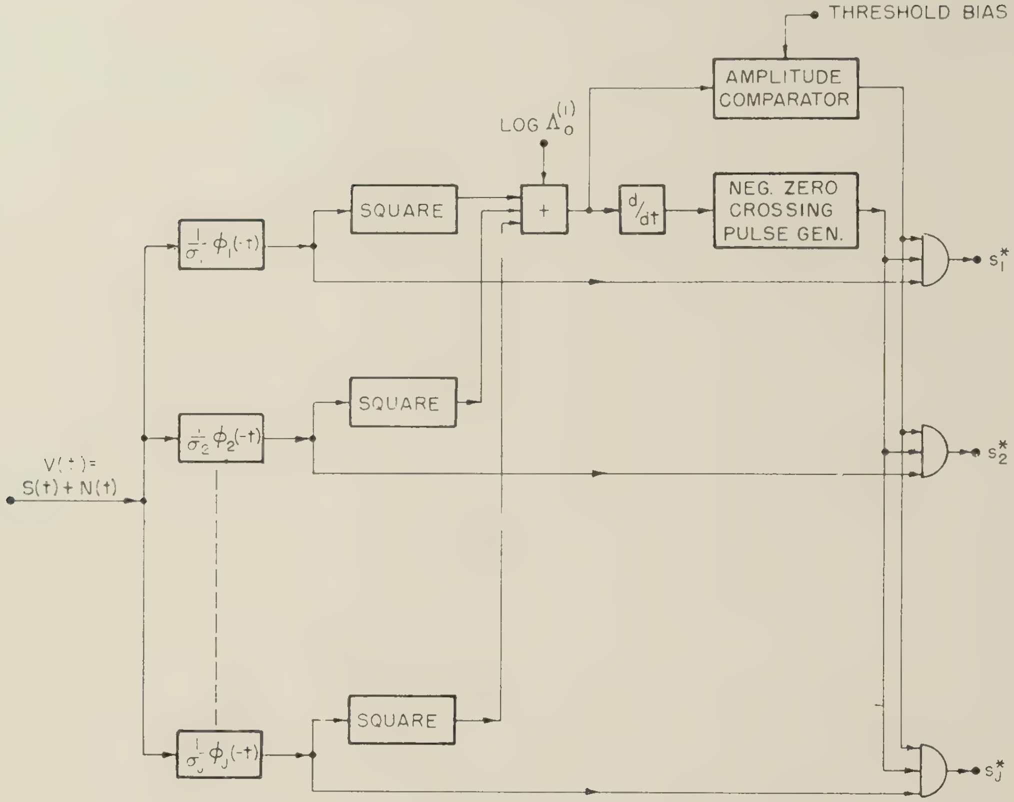


Fig. 1—Incoherent detection and estimation system, first J signal coefficients.

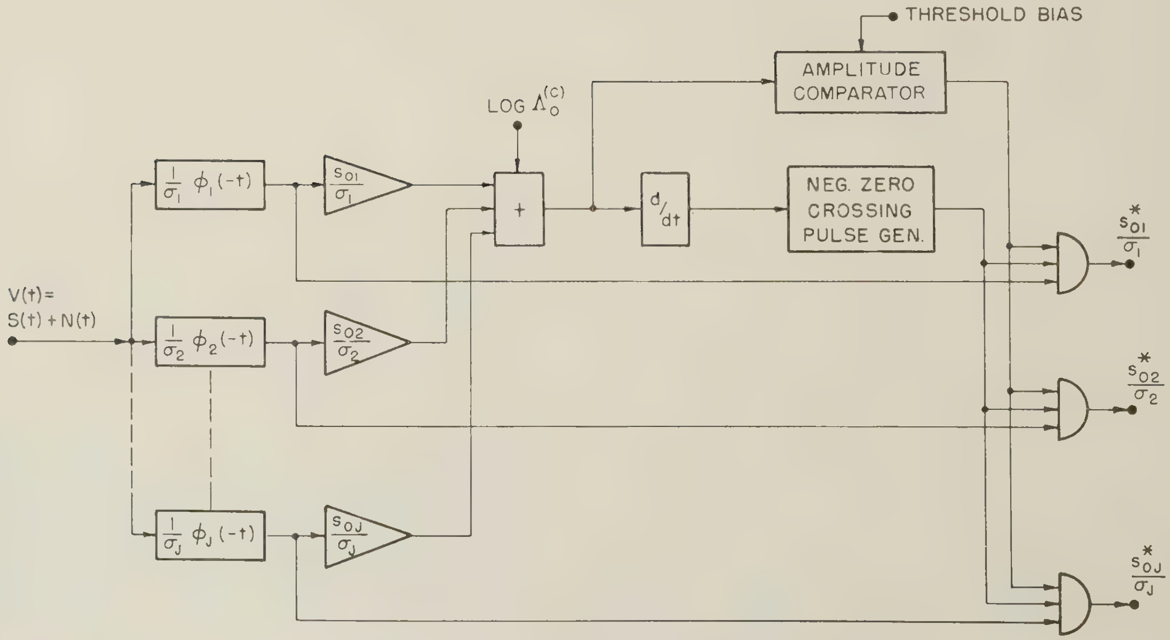


Fig. 2—Coherent detection and estimation, first J signal coefficients.

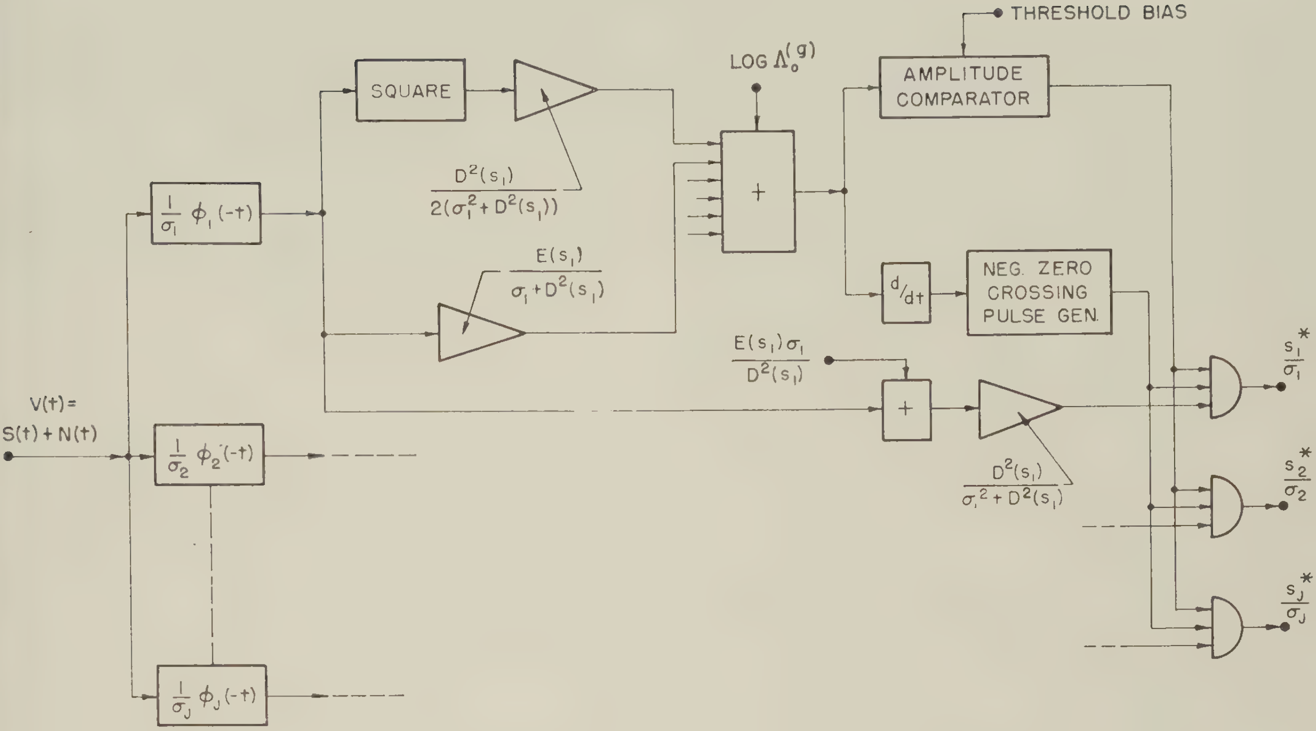


Fig. 3—Detection and estimation system for *a priori* normally distributed signals.

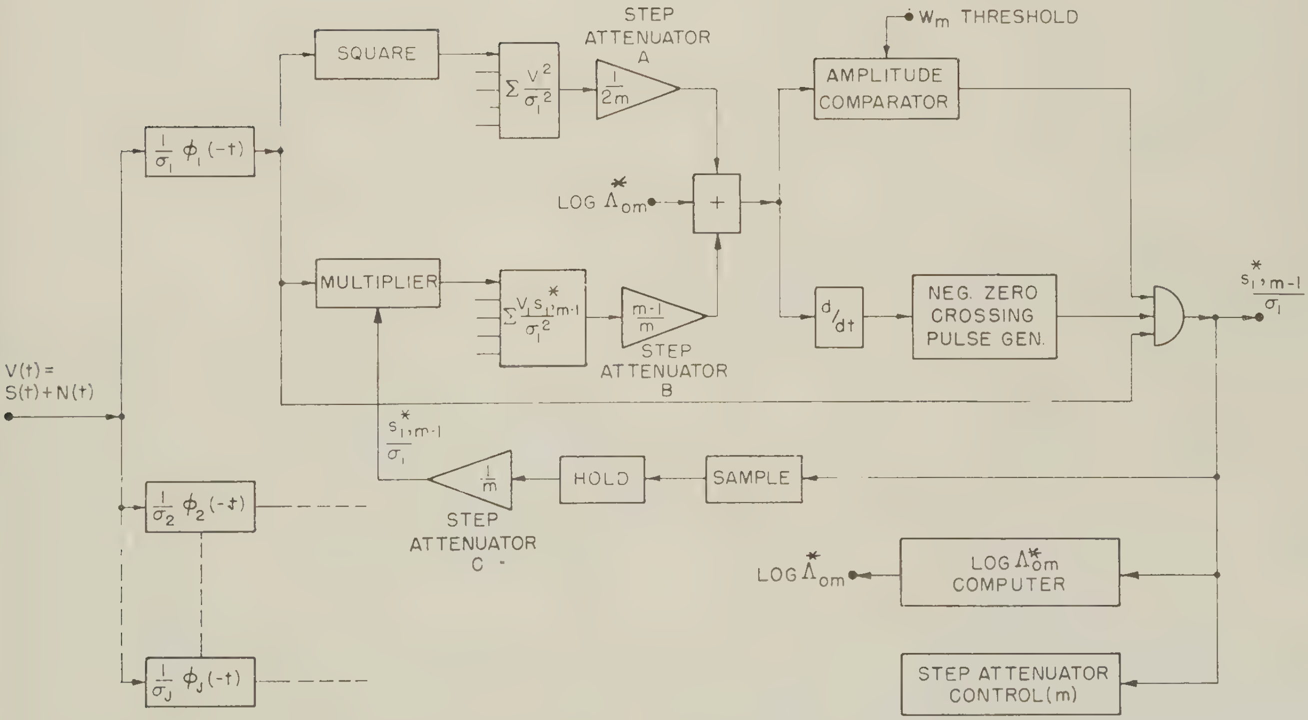


Fig. 4—Adaptive filter system for constant-amplitude pulse signals.

its initial incoherent form to the final coherent form. It does this by utilizing its accumulated estimates of the incoming signal in a prescribed fashion which is optimum. It is optimum because at each stage of adaptation, assuming the *a priori* signal distributions are correct, the system is set up to compute the log likelihood ratio and the signal parameter estimates called for in the previous section

The system operation can be outlined briefly. In the initial state the system is set up to receive in an optimum fashion any signal which may occur. All signals of duration less than T are assumed to be equally likely, and so the system initially takes the form of a square law detector followed by an integrator, (4) and (7). When a detection is made on the basis of the energy in the received data $V(t)$, an estimate of the signal parameters, (9), is made and used to modify the system structure. The modification is in accordance with the assumption that the received signal arises from a set of signals whose components are distributed in signal space in a Gaussian fashion. The expected values of the signal components are taken to be the measured parameters, and the variances are the background noise powers. The optimum system structure then must conform to that given in (17), resulting in a combination of linear and square law operations on the linearly filtered components of the incoming data. It maintains this general structure throughout the succeeding steps of adaptation. The system now detects a second pulse of the signal, and re-estimates the parameters of the signal. The estimation is now performed according to (19), for the distribution of signal parameters is now taken to be Gaussian instead of uniform. The optimum estimate is a weighted sum of the expected value of the parameter and the magnitude of the data at the read-out time. The weights themselves are determined by the distribution variances. The new estimates are used to readjust the system, still assuming a Gaussian distribution of signal parameters with means given by the estimates and variances half that of the background noise powers. This process then continues on with the third detection and parameter estimation, system readjustment, etc. It can continue until a given number of steps of adaptation are taken, or until some criterion of goodness of adaptation is satisfied.

Filter Operation and Structure, Fixed Amplitude Signals

The regimen of operation of the adaptive filter system (Fig. 4) is as follows:

1) At time $t = 0$, the filter is in a configuration for incoherent signal reception. The *a priori* distribution of signals in Ω is taken to be uniform, and only p, q and d, the expected duty ratio, are assumed known. The actual received signal, however, is $\mathbf{S}_0$. The J orthogonal filters have impulse responses $\phi_j(-t)$. Their outputs are squared and summed to yield the logarithm of the likelihood ratio $\log \Lambda_1^*$ for detection of the first pulse. Λ_m^* denotes the likelihood ratio computed by the system in its mth adaptive step. The detection threshold is set at W_1. The magnitude of W_1 is determined by the decision costs. The constant $\mu_1 = pd/1\text{-}pd$ and is part of $\log \Lambda_{01}^*$ [see (5)].

2) At $t = t_1$, $\log \Lambda_1^*$ exceeds W_1 and attains some maximum value. At this instant, the optimum estimates s_{j1}^*/σ_j are read out, stored and applied, via the multipliers, as filter gains to the filter outputs v_j/σ_j. The constant μ_1 is now adjusted to

$$\mu_2 = \frac{d/2}{1 - d/2}, \tag{20}$$

expressing a revaluation of the *a priori* probability from p to 1/2. The *a priori* distribution of signal is also altered from the uniform distribution to a Gauss distribution with mean $E(s_j) = s_{j1}^*$ for coefficient s_j and variance $D^2(s_j) = \sigma_j^2$. All coefficients are taken to be statistically independent. With this new *a priori* distribution, the system is modified in accordance with (17) to compute

$$\log \Lambda_2^*(t) = \log \Lambda_{02}^* + \tfrac{1}{4} \sum_{j=1}^{J} \frac{v_j^2(t)}{\sigma_j^2} + \tfrac{1}{2} \sum_{j=1}^{J} \frac{s_{j1}^* v_j(t)}{\sigma_j^2} \tag{21}$$

with

$$\log \Lambda_{02}^* = \log \left(\frac{d/2}{1 - d/2}\right) - \frac{J}{2} \log 2 - \tfrac{1}{4} \sum_{j=1}^{J} \frac{s_{j1}^{*2}}{\sigma_j^2}. \tag{22}$$

A new threshold W_2 is set automatically, again as a function of the decision costs involved. The system now has both incoherent and coherent contributions to $\log \Lambda_2^*$, the coherent output being weighted in accordance with the magnitude of the previous component estimates.

3) With the system operating as described above, a second pulse is detected at $t = t_2$ when $\log \Lambda_2^*$ reaches a maximum greater than W^2. The estimates s_{j2}^*/σ_j are read out, stored and applied as gains to the filter outputs. The estimators s_{j2}^* are obtained from (19).

$$s_{j2}^* = \frac{v_j(t_2)}{2} + \frac{s_{j1}^*}{2}. \tag{23}$$

The constant μ_2 is adjusted to $\mu_3 = (2d/3)/1 - 2d/3)$ as the *a priori* signal probability is now taken to be 2/3. A third threshold W_3 is set on $\log \Lambda_3^*$. A new computation is set up for $\log \Lambda_3^*$. It is obtained by taking as the *a priori* signal distribution one which is again Gaussian, with $E(s_j) = s_{j2}^*$ and $D^2(s_j) = D^2(s_{j2}^*) = \frac{1}{2}\sigma_j^2$, from (23). The signal coefficients are again assumed statistically independent. Then, referring to (17),

$$\log \Lambda_3^*(t) = \log \Lambda_{03}^* + \tfrac{1}{6} \sum_{j=1}^{J} \frac{v_j^2(t)}{\sigma_j^2} + \tfrac{2}{3} \sum_{j=1}^{J} \frac{s_{j2}^* v_j(t)}{\sigma_j^2}. \tag{24}$$

and

$$\log \Lambda_{03}^* = \log \left(\frac{2d/3}{1 - 2d/3}\right) + \frac{J}{2} \log \tfrac{2}{3} - \tfrac{1}{3} \sum_{j=1}^{J} \frac{s_{j2}^{*2}}{\sigma_j^2}. \tag{25}$$

) At $t = t_3$, a third pulse is detected and the process cribed above is repeated. The optimum coordinate mator for the third pulse is, by means of (19),

$$s^*_{j3} = \frac{v_j(t_3)}{3} + \tfrac{2}{3} s^*_{j2}. \tag{26}$$

) This process continues on as long as desired. For the reception, the log Λ^* computation is

$$\Lambda^*_m(t) = \log \Lambda^*_{0m} + \frac{1}{2m} \sum_{j=1}^{J} \frac{v_j^2(t)}{\sigma_j^2} + \frac{m-1}{m} \sum_{j=1}^{J} \frac{s^*_{j,m-1}\, v_j(t)}{\sigma_j^2}. \tag{27}$$

$$\Lambda^*_{0m} = \log \left[\frac{(m-1)\, d/m}{1 - (m-1)\, d/m} \right] - \frac{J}{2} \log \left(\frac{m-1}{m} \right) - \frac{m-1}{2m} \sum_{j=1}^{J} \frac{s^{*2}_{j,m-1}}{\sigma_j^2} \tag{28}$$

d

$$s^*_{jm} = \frac{v_j(t_m)}{m} + \frac{(m-1)}{m} s^*_{j(m-1)} = \frac{1}{m} \sum_{\mu=1}^{m} v_j(t_\mu). \tag{29}$$

s easy to see that as m becomes infinite,

$$\underset{m\to\infty}{\mathrm{l.i.m.}}\ s^*_{jm} = s_{0j} \tag{30}$$

l, as a result,

$$\text{m. } \log \Lambda^*_m(t) = \log \left(\frac{d}{1-d} \right) - \tfrac{1}{2} \sum_{j=1}^{J} \frac{s_{0j}^2}{\sigma_j^2} + \sum_{j=1}^{J} + \frac{s_{0j} v_j(t)}{\sigma_j^2}. \tag{31}$$

is is identical to (15) when $\mu = d/(1-d)$, so that the cess of adaptation described results in the system verging to the configuration of a filter matched to the eived signal $\mathbf{S}_0$.

cussion of the Adaptive Process

The description of the adaptive process and derivation the functions $\log \Lambda^*_m$ and s^*_{jm} have been carried out on assumption that there is no error involved in estiting the epoch of the pulses. This, of course, is not true. tead of $\log \Lambda^*_m$ reaching its maximum at $t = t_m$, which vould if there were no noise, the maximum is actually ched at time t^*_m. As a result, the signal estimates (t^*_m) have two sources of error, the noise amplitude at nd the epoch error $e_m = t_m - t^*_m$. The statistics of the er error are quite difficult to obtain, but are a func- of the shape of the signal and its signal-to-noise ratio le those of the former are Gaussian and constant for tionary Gauss noise. For weak signals, the component mates will tend to be large and induce erratic behavior the adaptive system. It appears that there may be a minimum signal strength which the system can adapt to, and this also may be a function of signal waveform. Signals above this minimum will be successfully adapted to, and at a rate which decreases as signal strength increases. No calculations of this minimum "adaptable" signal have been made. During the reception of a signal of this type, as the number of pulses received increase, and coherent operation is approached, the epoch estimate errors will decrease. The optimum estimator which has the form given by (29), for no epoch estimation error, will have instead the form

$$s^*_{jm} = \sum_{\mu=?}^{m} \alpha_\mu v_j(t^*_\mu), \tag{32}$$

where

$$\alpha_\mu \geq \alpha_{\mu-1} \qquad \mu = 1, 2, \cdots, m.$$

That is, the recent estimates will be weighted more heavily than the more remote ones. The weighting parameters will be a function of the signal waveform. Since these are unknown initially, they could only be obtained by another process of estimation relying upon the preceding estimates.

The detection thresholds W_m are functions of the costs of decision errors, as mentioned previously. One convenient choice of threshold is that which maintains a fixed false alarm rate. The threshold is then calculated from the statistics of the background noise. Such a calculation is simple for either the incoherent or coherent detector, for in these cases the statistics are either chi-square or Gaussian. For the intermediate case where the noise is a combination of the two, (27) with $m > 1$, no tables exist. The threshold setting on $\log \Lambda^*_m$ is dependent upon the signal parameter estimates. The exact form of the dependency has not yet been obtained. Approximation techniques must be resorted to.[5] As a result, no rates of occurrence of false alarms have been calculated.

Filter Structure and Operation, Variable Amplitude Signals

The adaptive process as described above is adequate for the reception of an unknown signal of fixed shape and amplitude. If the amplitude of the signal fluctuates while the shape remains constant, the adaptive process described above is still applicable, provided the signal estimates are reduced to a normalized form $\bar{s}_{j,m}$ before being applied as filter gains $\bar{s}_{j,m}/\sigma_j$ to the outputs of the orthogonal filters. We have here

$$\frac{\bar{s}_{j,m}}{\sigma_j} = \frac{s^*_{j,m}/\sigma_j}{\left[\sum_{k=1}^{J} \left(\frac{s^{*2}_{k,m}}{\sigma_k} \right) \right]^{1/2}}. \tag{33}$$

The purpose involved in using the normalized estimates is to eliminate the effects of the amplitude fluctuations in the signal and to simplify somewhat the design of a work-

[5] *Ibid.*, ch. V.

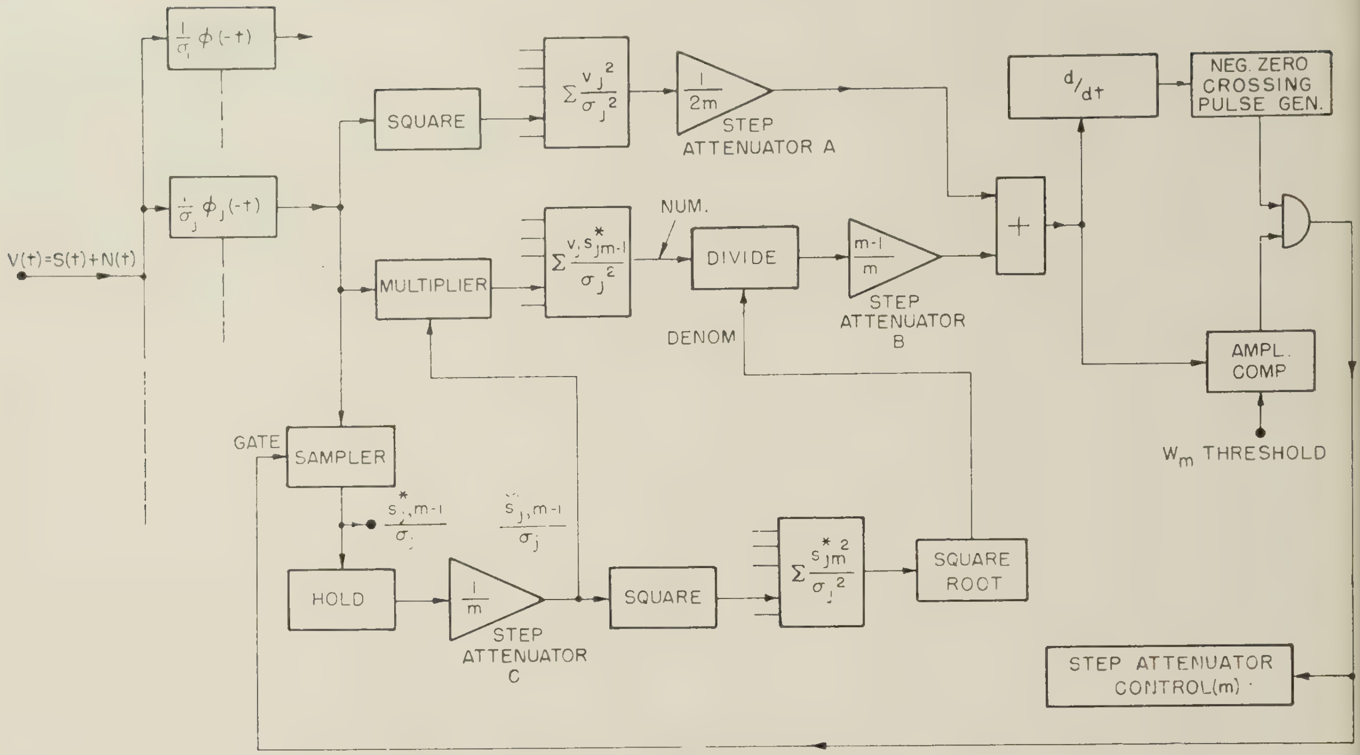


Fig. 5—Normalized adaptive filter system for detecting pulse signals with fluctuating amplitudes.

able adaptive filter. The structure of an adaptive filter using normalized component estimates is shown in Fig. 5.

It is to be pointed out here that this filter is no longer optimum either for detecting signals of fixed or fluctuating amplitudes. In the former case, Park [6] has discussed the structure of an optimum system which estimates the normalized components directly, rather than normalizing the estimates as is done here. In the latter, fluctuating-signal situation, design of an optimum system requires knowledge of the statistical nature of the signal fluctuations.

Signal detection and epoch estimation is now based upon the quantity $\log \tilde{\Lambda}_m$, the log of the likelihood ratio for normalized estimation. This is given by a slight modification of (27) and (28):

$$\log \tilde{\Lambda}_m(t) = \log \tilde{\Lambda}_{0m} + \frac{1}{2m}\sum_{j=1}^{J}\frac{v_j^2(t)}{\sigma_j^2} + \frac{m-1}{m}\sum_{j=1}^{J}\frac{\tilde{s}_{j,m-1}v_j(t)}{\sigma_j^2} \quad (34)$$

$$\log \tilde{\Lambda}_{0m} = \log\left[\frac{\left(\frac{m-1}{m}\right)d}{1-\left(\frac{m-1}{m}\right)d}\right] - \frac{J}{2}\log\left(\frac{m-1}{m}\right) - \frac{m-1}{2m}\sum_{j=1}^{J}\left(\frac{\tilde{s}_{j,m-1}}{\sigma_j}\right)^2. \quad (35)$$

This last expression can be simplified by noting that the summation in the last term on the right-hand side is one identically. Then

$$\log \tilde{\Lambda}_{0m} = \log\left[\frac{\left(\frac{m-1}{m}\right)d}{1-\left(\frac{m-1}{m}\right)d}\right] - \frac{J}{2}\log\left(\frac{m-1}{m}\right) - \frac{m-1}{2m}. \quad (36)$$

There is some simplification found in the use of $\log \tilde{\Lambda}_m(t)$ instead of $\log \Lambda_m^*(t)$, for the bias term $\log \tilde{\Lambda}_{0m}$ is a constant independent of the estimates of the received signal. The functional form of the fluctuating component is, however, a function of $\tilde{\mathbf{S}}_{m-1}$ and quite difficult to derive.

The probability density functions for the normalized estimates are quite complicated. When the signal-to-noise ratio is large, Park[6] has found that

$$E\left(\frac{\tilde{s}_{j,m-1}}{\sigma_j}\right) \approx \frac{s_{nj}/\sigma_j}{\left[\sum_{k=1}^{J}\left(\frac{s_{nk}}{\sigma_k}\right)^2\right]^{1/2}} \quad (37)$$

and

$$D^2\left(\frac{\tilde{s}_{j,m-1}}{\sigma_j}\right) = \frac{1}{m-1} \times \text{cons.}, \quad (38)$$

[6] See [6], ch. IV, sect. D.

ere s_{nj} is the normalized jth component of the received
nal. Then it is true that

$$\underset{m\to\infty}{\text{l.i.m.}}\left(\frac{\tilde{s}_{j,m-1}}{\sigma_j}\right) = \frac{1}{\sqrt{C_J}}\frac{s_{nj}}{\sigma_j}, \tag{39}$$

h

$$C_J = \sum_{k=1}^{J}\left(\frac{s_{nk}}{\sigma_k}\right)^2.$$

om this, (34), and (35) it can be seen that

$$\text{m.}_{\to\infty} \log \tilde{\Lambda}_m(t)$$

$$= \log d - \tfrac{1}{2} + \frac{1}{\sqrt{C_J}}\sum_{j=1}^{J}\frac{s_{nj}v_j(t)}{\sigma_j^2}. \tag{40}$$

us, when the signal-to-noise ratio is large enough for) and (38) to apply, the adaptive system for reception an amplitude fluctuating signal will converge to a filter ich is matched to a signal whose components are $/\sqrt{C_J}$.

Synthesis of an Adaptive System (Normalized)

An attempt was made to synthesize an adaptive system means of an analog computer installation. The back- ound noise was white, and the orthogonal filters were osen to be the Laguerre functions. The system was signed to work in real time. Considerable difficulties re encountered in achieving satisfactory results, owing inly to deficiencies in the available analog computer uipment. A description of the system is given in Chapter of [5].

The unsatisfactory performance of the analog computer simulating the adaptive filter during experimental ts led to the investigation of methods of simulating e adaptive system with a general purpose digital com- ter. It was found that such a simulation was quite sible when the computer was used to represent not ly the filter system, but also the signal source and the erfering noise. This type of simulation is somewhat ore abstract than that performed by the analog com- ter, in that there is no use of physical signal or noise rces, or of electrical filter networks. Only the math- atical properties of the sources and filters are used. so, the experiment no longer takes place in real time, ce the generation in the computer of signal and noise ocesses is required to await the completion of previous tection and estimation computations. There is, however, great advantage in using the digital computer. The per- mance of the computer is known exactly, so that all ects observed are due only to the mathematical proper- s of the adaptive system and not to the form of its ysical realization. There is never confusion as to whether observed effect is an inherent result of the adaptive ocess or of a defect in the hardware used in the nthesis.

It is desirable to make some changes in the actual system to be synthesized when a digital computer is used instead of an analog computer. These changes involve primarily the use of a different set of orthogonal filters to represent the signal and noise processes. In the analog computer, Laguerre functions are easy to employ; in the digital computer, where a sampled-data type of operaton occurs, the cardinal functions [(sin x)/x] are appropriate. To be more precise, the digital computer can generate samples of the signal and noise processes at discrete points in time, and can only perform calculations based on these discrete samples. It is convenient to assume that the signal-plus-noise samples are spaced equally in time, with a separation of τ_0 seconds. Then if the noise is white, and bandwidth limited to the frequency interval $(-1/2\tau_0, 1/2\tau_0)$, the autocorrelation function of the noise is

$$K_N(t, u) = N\frac{\sin 2\pi f_0(t-u)}{2\pi f_0(t-u)}, \quad \text{where} \quad f_0 = \frac{1}{2\tau_0}, \tag{41}$$

and the integral equation (37) becomes

$$\int_{-\infty}^{\infty} N\frac{\sin 2\pi f_0(t-u)}{2\pi f_0(t-u)}\phi_j(u)\,du = \sigma_j^2\phi_j(t). \tag{42}$$

This is satisfied by the set of cardinal functions defined by

$$\phi_j(t) = \frac{\sin 2\pi f_0(t-j\tau_0)}{2\pi f_0(t-j\tau_0)} \quad j = 0, \pm 1, \pm 2, \pm 3 \cdots \tag{43}$$

The $\phi_j(t)$ are orthogonal over the interval $(-\infty, \infty)$. The characteristic values of the integral equation are equal and given by $\sigma_j^2 = N_0 = N/2f_0$, the average per cps.

The noise $N(t)$ can now be written as

$$N(t) = \sum_{j=-\infty}^{\infty} n_j \frac{\sin 2\pi f_0(t-j\tau_0)}{2\pi f_0(t-j\tau_0)} \tag{44}$$

and

$$N(j\tau_0) = n_j. \tag{45}$$

The values of $N(t)$ at the sampling instants $j\tau_0$ are the coefficients of the cardinal functions in the orthogonal expansion of $N(t)$, $(-\infty < t < \infty)$.

In any data processing task involving a digital computer we do not have sufficient capacity to permit storage of data taken an infinite time ago. Consequently, we are interested in a system which stores only the most recent M sample values. From (45), it can be seen that a sequence of M samples of the noise taken τ_0 seconds apart is equivalent to the outputs of M cardinal function filters. If the past $M - 1$ samples are stored and processed together with the current sample, we have a filter system which is a member of the class of filter systems discussed previously. The system operation of processing the M most-recent samples simultaneously can also be seen to be similar to the filter system of Fig. 5, when ϕ_1 is a cardinal function filter, ϕ_2 is replaced by a τ_0 second delay line in cascade with another ϕ_1 cardinal function filter, and ϕ_j is replaced by a $j\tau_0$ second delay line in cascade with a ϕ_1

filter. The system modification is illustrated in Fig. 6. The fact that the cardinal functions cannot be synthesized exactly is not important, since the filter output is what is of importance, and this is obtained by the sampling procedure.

There is a difference between the actual sampling process and the extraction of orthogonal coefficients by means of the system of Fig. 6. The sampling process will yield values for the logarithm of the likelihood ratio only at the times corresponding to the sampling times. No continuous computation of log Λ can be performed. Consequently, the optimum read-out time is constrained to be at one of the sampling times, and some loss is found in the performance of the sampled-data adaptive system which does not occur in the continuous data processing system. This loss, though not determined, would not seriously degrade the performance of the system. The reason is that for a short duration signal, where epoch error is most important, the signal amplitude must be high for initial detection. Consequently, little error will occur in the initial parameter estimation, and this will decrease with subsequent estimates. For relatively long duration signals, the log of the likelihood ratio will tend to have a broad maximum. Epoch estimation errors in this situation will, therefore, not cause great errors in the waveform estimates. Nevertheless, the adaptive mechanism in both systems is the same.

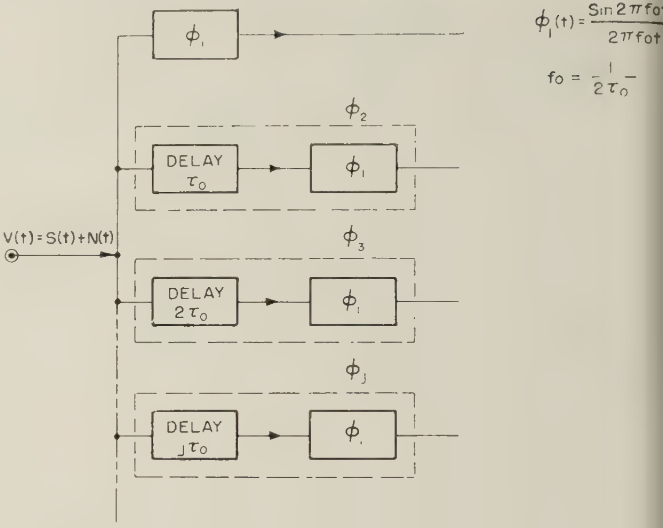


Fig. 6—An orthogonal filter system using cardinal functions (see Fig. 5).

The Simulation Program

The UNIVAC scientific digital computer at The Johns Hopkins University, Applied Physics Laboratory, was employed in this simulation study. The entire program was written in APT (Automatic Program Translator) Code.

1) Number of Samples (M): The number of signal-plus-noise samples (corresponding to the orthogonal components) was chosen arbitrarily to be 10. For the particular signal chosen (see below), this guarantees that the signal can be represented without error by this set of orthogonal components, and that, in the limit, the system will adapt to the signal exactly. In the more general situation, the set of orthogonal components used to represent the signal, (2) and (3), and the number of their corresponding filters employed in the system synthesis are vital to the speed of convergence of the adaptive system to the received signal. For signals of equal energy, adaptation will be most rapid for those whose waveforms are most closely approximated by the set of filters in use.

2) Signal: The only signal pulse used in this program was a square wave which was represented in sampled data form by a sequence of 5 samples of equal magnitude. The duty ratio was taken to be 0.05 so that there were 95 consecutive samples of pure noise followed by 5 of signal plus noise.

3) Noise: The noise samples had zero mean and unit variance. They were obtained from a computer subroutine which consisted of the generation and addition of twelve random numbers uniformly distributed between 0 and 1, subtraction of the constant 6 and division by twelve. The sum so obtained is very nearly Gaussian.

4) Filter Operations: The filter operations of squaring, summing, amplitude comparison, and estimate normalization were performed substantially as in the analog computation. Detection thresholds were the same as before. The optimum read-out time computation was somewhat different. Here, the current value of log $\tilde{\Lambda}_m$ was computed and stored if it exceeded both the amplitude comparison threshold and the previous high value of log $\tilde{\Lambda}_m$. Also stored temporarily were the ten sample values of the input which produced this value of log $\tilde{\Lambda}_m$. A running count was made of the number of samples taken from the time log $\tilde{\Lambda}_m$ first exceeded the amplitude threshold. When this count reached twenty, the temporarily-stored sample values were transferred to the signal estimate accumulators, and the log $\tilde{\Lambda}_m$ address was cleared to zero. Detection thresholds were changed at the end of every detection as was the mixing ratio for coherent and incoherent filter outputs. To prevent the computer from running for excessively long times without a successful detection, a count was kept of the number of signal pulses generated during each run. A run was terminated after 20 signal pulses were generated, whether or not there were 10 signal (or noise) pulse detections.

5) Tests: Ten detection runs were made at each of four values of signal-to-noise ratio. Each run consisted of a sequence of ten detections. The signal-to-noise ratios were calculated from the definition

$$S/N_0 = \frac{\text{signal pulse energy}}{\text{av. noise power/cps}}.$$

The signal pulse energy is $5S^2\tau_0$, where S is the magnitude

[7] See [7], pp. 244–245.

each of the five signal samples and τ_0 is the sampling erval. The average noise power is 1 and the noise bandlth is $1/\tau_0$. Then,

$$S/N = \frac{5S^2\tau_0}{\tau_0} = 5S^2. \quad (46)$$

e values of S/N_0 tested were 100, 25, 16, and 9. The rresponding db values for S/N_0 were 20, 14, 12, and 9.5.

5) *Test Results*: The data obtained from the detection ns at each particular signal-to-noise ratio were processed yield:

a) average values of signal component estimates;
b) estimates of the variances of the signal component estimates;
c) the fraction of the estimated signal-waveform energy contained in those estimate components corresponding to the nonzero signal-waveform components.

These data are shown plotted in Figs. 7–9. Smooth rves have been drawn to fit these points, the assumpn being that if sufficient runs had been taken, the perimental points would fall very close to these curves. For the runs made at $S/N_0 = 9$, only two were cometed (ten detections) before 20 signal pulses had been ansmitted. (The results for these runs are inadequate d are not shown here.) Nine detection runs were cometed at $S/N_0 = 16$ and at the higher values of S/N_0 all ere completed.

The average of the signal component estimates and eir variances included all the component estimates gether since the individual nonzero signal components ere chosen to be equal. This procedure washes out the fects of the estimation process on the estimates of the dividual signal components. However, from the data tained, there did not appear to be a significant variation the average of the estimates of the individual comnents or in their variances. An investigation of effects the individual component estimates would have been ore appropriate if the amount of data taken had been uch greater. As it was, the data taken were adequate to veal only the grosser aspects of the adaptive process.

iscussion of Results

The value for each normalized signal component is $\sqrt{5} = 0.447$. It can be seen that the average comnent estimates started low for all S/N_0 values and proached this value as the number of detections ineased. In the weaker signal situations, both the initial d final estimates are poorer (lower). The five final mponent estimates corresponding to the nonzero signal mponents contain over 90 per cent of the total estimated ergy for all $S/N_0 \geq 16$.

In Table I are shown the per unit error in the erage value of the signal component estimates, the imate variances, and the per unit energy content of the e nonzero components of the estimated signal, all at e end of the adaption process.

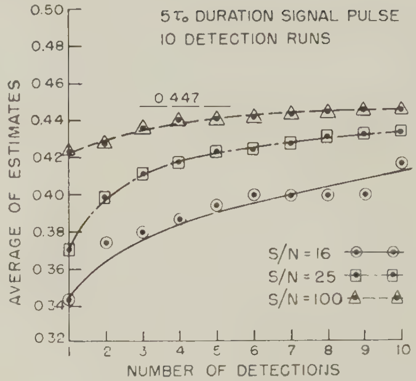


Fig. 7—Adaptive filter digital simulation results.

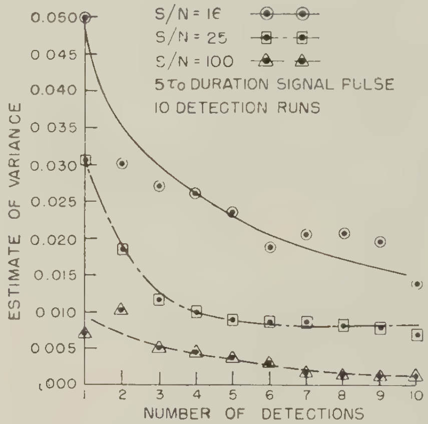


Fig. 8—Adaptive filter digital simulation results.

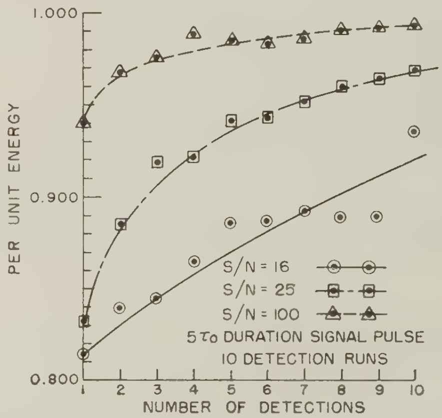


Fig. 9—Adaptive filter digital simulation results.

TABLE I

	S/N_0		
	16	25	100
Average Estimate Error	−0.07	−0.03	−0.01
Variance of Estimate	0.014	0.007	0.0015
Energy Content of Nonzero Components	0.94	0.97	0.992

From the table it may be seen that the estimate variances increase almost in inverse proportion to S/N_0, and in direct proportion to noise power. This effect is to some extent predicted by (38) which can be rewritten as

$$D^2(s^*_{i,m-1}) = \frac{\sigma_i^2}{m-1}. \quad (47)$$

That is, the variance of the unnormalized signal estimate is proportional to the noise power. The variance of the normalized signal estimate is more complex, as mentioned before, but tends to this result when the signal-to-noise ratio is large, which is the situation here.

The average estimates are all consistently low throughout the tests, although they do tend to approach the true signal component values more quickly as S/N_0 increases. This effect was not predicted in the theoretical analysis, mainly because of the difficulties encountered in working with the estimators of normalized signal components. It is apparent, however, that when the signal consists of a group of equal nonzero components and another group of zero components, the average of the normalized estimates of the nonzero group must always be less than the true average of this group. As the magnitudes of the normalized estimates of the zero components of the signal decrease, the average of the estimates of the nonzero components will increase to the true average value. This is the situation which occurred in these tests. There were five equal nonzero signal components and five zero signal components.

With regard to the concentration of energy of the estimated signal in the components corresponding to the nonzero signal components, it can be seen that the adaptive system has substantially succeeded in recognizing the fact that five of the estimated components are zero or very small. Thus, although the component estimates are in error by 7 per cent or more for low values of S/N_0, a good degree of match has still been achieved.

A more thorough study of the behavior of the system at low values of S/N_0 and for different signal shapes would have been desirable. This would have required setting higher detection thresholds to lower the false alarm rate and would have increased the required machine time.

Conclusions

The results of the digital computer simulation demonstrate that system adaptation to the received signal waveform takes place for signal-to-noise ratios as low as 16 (12 db). Below this level, the amount of data obtained is insufficient to yield any significant conclusions concerning adaptation. It would appear that the adaptive system would perform satisfactorily at significantly lower values of signal-to-noise ratios, perhaps as low as 6 db, if the false alarm probability were decreased by several orders of magnitude. This would, of course, also decrease the probability of detection, but would still yield a system superior in detection performance to a nonadapting incoherent system.

The statement asserting superiority of the adaptive system to the incoherent system assumes, of course, the validity of the assumptions concerning the uniformity of the *a priori* signal distribution, and the relatively slow variations in received signal waveform from pulse to pulse. It also assumes that the use of a small number of orthogonal components to represent the signal waveform will not result in a significant waveform approximation error. These assumptions have not been examined closely. As a result, questions remain open as to whether the actual increase in signal detectability and the reduction in error rates of an adaptive system of this type justify the increase in the structural complexity of such a system over the more conventional invariant optimum systems. They can probably only be answered when a quite specific problem in signal detection is considered.

The present study has demonstrated that it is possible to design and construct a signal detection system which is capable of adapting its structure to match that of an incoming signal. The usefulness of this type of system in various signal detection problems can be great. Detection and analysis of signals of unknown (*a priori*) characteristics is one possible application. Another is the tracking of known but slowly and randomly varying signals.

Acknowledgment

The author wishes to express his thanks to Profs. W. H. Huggins and W. C. Gore, and to Dr. J. H. Park, Jr., all of whom contributed in good measure to the pursuit of this investigation.

References

[1] N. Wiener, "The Extrapolation, Interpolation, and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, New York; 1949.
[2] R. Price and P. E. Green, Jr., "A communication technique for multipath channels," Proc. IRE, vol. 46, pp. 555–570; March, 1958.
[3] D. G. Brennan, "Linear diversity combining techniques," Proc. IRE, vol. 47, pp. 1075–1102; June, 1959.
[4] D. Middleton and D. Van Meter, "Detection and extraction of signals in noise from the point of view of statistical decision theory, parts I and II," *J. Soc. Ind. Appl. Math.*, vol. 3, pp. 192–253, December, 1955; and vol. 4, pp. 86–119, June, 1956.
[5] E. M. Glaser, "Signal Detection by Adaptive Filters," Rad. Lab., The Johns Hopkins University, Baltimore, Md., Tech. Rept. AF-75; April, 1960 (Unclassified).
[6] J. H. Park, Jr., "Statistical Estimation of Normalized Linear Signal Parameters," Rad. Lab., The Johns Hopkins University, Baltimore, Md., Tech. Rept. AF-72; December, 1959 (Unclassified).
[7] H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J.; 1946.
[8] D. Middleton, "An Introduction to Statistical Communication Theory," McGraw-Hill Book Co., Inc., New York, N. Y.; 1958.

The Coding of Pictorial Data*

JOSEPH S. WHOLEY†

ummary—We are concerned with the problem of designing efficient general method of coding two-level pictorial data. h exact and approximate coding techniques are illustrated. pilot experiment is presented, in which a digital computer used to realize two-dimensional "predictive coding." Although resulting compression was not great, there are reasons for eving that this procedure would be more successful with realistic orial data.

urther experiments, which made use of approximation methods described. These methods arose from the application of pattern gnition theory to the present problem. Their use, either in-endently or prior to predictive coding, yielded compression nificantly greater than that attained by predictive coding alone.

I. Introduction

WE ARE INTERESTED in coding m-by-n matrices of black and white elements which represent pictures, maps, or other "meaningful" tterns. Since only those matrices having some definite ucture will be coded, we expect that binary codes sub-ntially shorter than mn bits will be obtainable for each the patterns to be coded. A goal might be to take, say, 0-by-500 matrices of black and white elements which present pictures and assign to them codes which average, y, less than 25,000 bits. Since we make the approxi-ation that the pictures to be considered are equiprobable, is would imply that there are no more than $2^{25,000}$ dif-ent meaningful pictures representable on a 500-by-500 atrix. We describe first an exact coding process which has en performed by a computer and second our experiments th approximation methods, for which the additional ograms were not actually constructed, but only simu-ed.

II. Predictive Coding: An Exact Coding Process

A method[1,2] of exact coding of pictorial data has been tained as an extension of the work of Elias[3,4] and of hreiber.[5,6] The method used is predictive coding, which involves the following steps (the first two of which are thought of as steps in a learning process, rather than as part of the predictive coding process itself):

1) A survey of pictures representative of the class to be coded in any particular application is made to determine the relative frequency with which each of the possible "neighborhood" patterns X is followed by a black element. A neighborhood of a matrix element y is some convenient selection (based on available storage and number of pictures surveyed) of the elements which precede y in the matrix, when the matrix is scanned in some standard fashion, *e.g.*, left to right, from top to bottom. Since we are concerned with a matrix rather than a sequence, a two-dimensional neighborhood is indicated as the best choice in general.

In the experiments[1] so far performed, a twelve-element neighborhood X was used:

$$\begin{matrix} X_1 & X_2 & X_3 & X_4 & X_5 \\ X_6 & X_7 & X_8 & X_9 & X_{10} \\ X_{11} & X_{12} & y. & & \end{matrix}$$

Statistics were gathered on the number of times each of the 2^{12} different neighborhoods of black and white elements was followed by $y = B$ and the number of times by $y = W$.

Examples:

$$\text{a)}\quad \begin{matrix} B & B & B & B & B \\ B & B & B & B & B \\ W & W & y. & & \end{matrix}$$

A white element would usually occur in the y position, following *this* neighborhood.

$$\text{b)}\quad \begin{matrix} W & W & B & W & W \\ W & W & B & W & W \\ W & W & y. & & \end{matrix}$$

A black element would usually occur in the y position, following *this* neighborhood.

A Datatron 205 computer required about thirty minutes to compile neighborhood statistics for 7000-element matrices.

2) For each of the different neighborhoods, a unique prediction y is determined as the color which is most likely to follow the neighborhood (on the basis of the statistical survey). We thus obtain a "prediction function," a table in which each of the possible neighborhoods "predicts" (*i.e.*, is paired with) the color of y which is most likely to follow it.

* Received by the PGIT, June 23, 1960. The work reported re was done for Wright Air Dev. Div., Air Res. and Dev. Com-nd, under Air Force Contract No. AF 33(616)-5589.

† Melpar, Inc., Appl. Sci. Div., Watertown, Mass.

[1] "Flight Display and Flight Control Integration," Res. Dept., lbar, Inc., Watertown, Mass., AF 33(616)-5589; 1959.

[2] R. E. Wernikoff, "On the Efficient Representation of Pictorial ta," Appl. Sci. Div., Melpar, Inc., Watertown, Mass., Res. pt. No. 59/1; 1959.

[3] P. Elias, "Predictive Coding," Ph.D. dissertation, Harvard iv., Cambridge, Mass.; 1950.

[4] P. Elias, "Predictive coding," IRE Trans. on Information eory, vol. IT-1, pp. 16–33; March, 1955.

[5] W. F. Schreiber, "Probability Distributions of Television nals," Ph.D. dissertation, Harvard Univ., Cambridge, Mass.; 52.

[6] W. F. Schreiber, "The measurement of third-order probability tributions of television signals," IRE Trans. on Information eory, vol. IT-2, pp. 94–105; September, 1956.

Examples:

$$\text{a)}\quad \left.\begin{matrix} B & B & B & B & B \\ B & B & B & B & B \\ W & W & y & & \end{matrix}\right\} \Rightarrow y = W.$$

$$\text{b)}\quad \left.\begin{matrix} W & W & B & W & W \\ W & W & B & W & W \\ W & W & y & & \end{matrix}\right\} \Rightarrow y = B.$$

3) Now each matrix representing a picture is paired (in a one-to-one fashion) with an "error matrix" which indicates the elements y of the picture matrix for which the prediction function gives an incorrect value. In order to make possible a uniform prediction procedure, the picture is treated as if it were surrounded by a white border two elements wide. For each element y in the picture matrix, the actual color of y is compared with the color predicted by the function above, on the basis of the neighborhood of y. If the colors are the same, then a 0 is stored in the corresponding position in the error matrix; otherwise a 1 is stored there. (The computer performed this operation also in about thirty minutes.)

4) The error matrix (which under ideal conditions should contain a small percentage of 1's) is now coded by some process like run-length coding (which gives the number of 0's between successive pairs of 1's). Elias[4] has shown that, assuming uncorrelated error terms, a run-length code can always be found which codes the error terms efficiently: the average compression coefficient (code length divided by the number of elements in the error matrix) will not be much greater than the entropy, or optimum compression coefficient, H (for a binary array of uncorrelated terms in which the probability of one of the symbols is p, $H \doteq -[p \log_2 p + (1-p) \log_2 (1-p)]$).

5) To reverse the process and obtain a picture from its code number, we first obtain the error matrix from the run-length code. Then, the output of the prediction function for each successive element of a new picture matrix is compared with the corresponding element of the error matrix. The color predicted is entered in the new matrix if the corresponding element of the error matrix is a 0; if the corresponding element is a 1, the other color is entered. The completed matrix thus represents the desired picture.

III. Results of a Predictive Coding Experiment[1]

Our experiment was made on weather maps, in view of military requirements for efficient storage and transmission of information of this type. Figures representing sets of isobars from ten maps were traced on 70-by-100 matrices, an element of the matrix being counted as black if crossed by an isobar. Because of the jaggedness of the figures thus obtained, really spectacular results could not be expected from predictive coding. On the other hand, if the jaggedness is taken as representing the effects of noise in a full-scale pictorial data processing system, lack of success can be considered an illustration of Graham's[7] observation on the vulnerability of a predictive coding system (or indeed any exact coding system) to the effects of noise which is embedded in the original data.

Error matrices were obtained for each of the maps, using a prediction function defined on the basis of a statistical survey of the whole group of ten maps. Each of the error matrices was then coded by run-length coding. Figs. 1 and 2 are computer printouts of one of the maps and the corresponding error matrix. The results for the group of ten maps are summarized in Table I.

As was mentioned in Section II, 4), the optimum compression coefficient for an uncorrelated array of zeros and ones, where the probability of a one is 5.5 per cent, is $H = -[.055 \log_2 .055 + .945 \log_2 .945] = 0.31$.

The average compression obtained by predictive coding was thus rather unimpressive. Greater compression (*i.e.*, a smaller average compression coefficient) is expected for the high-resolution data which might be required in a military or commercial display device, since the high-resolution data would tend to be smoother and the proportion of errors, therefore, to be smaller. The following factors, on the other hand, will limit the compression obtainable by predictive coding:

1) The effects of noise and other irregularities introduced in representing the original (continuous) picture on a discrete matrix are limiting factors.
2) The inefficiency resulting from coding almost indistinguishable pictures with different (and therefore, on the average, longer) code numbers is also a limiting factor.
3) The "global" rather than local character of many of the constraints present in meaningful pictures limits the compression obtainable by predictive coding (Youngblood[8] argues that no coding scheme based on local operations is likely to be successful). Predictive coding will not take advantage of these "global" constraints, since the largest neighborhood that can profitably be considered (because of memory limitations) is not much larger than the twelve-element one which has been described.
4) Even supposing that, for 500-by-500 matrices, errors could be kept down to 2 per cent, an average code length shorter than 35,000 bits could not be achieved ($35{,}000 \doteq 250{,}000H$, where $H = -[.02 \log_2 .02 + .98 \log_2 .98] \doteq 0.14$).
5) If letters, numbers, and other fine detail were added to the picture, code lengths would be tremendously increased, although there must obviously be an efficient way to code alphanumeric and other familiar data efficiently.

[7] R. E. Graham, "Communication theory applied to television coding," *Acta Electronica*, vol. 2, 1–2, pp. 333–343; 1957–1958.

[8] W. A. Youngblood, "Estimation of the Channel Capacity Required for Picture Transmission," Sc.D. dissertation, Mass. Inst. Tech. Res. Lab. of Electronics, Cambridge, Mass.; 1958.

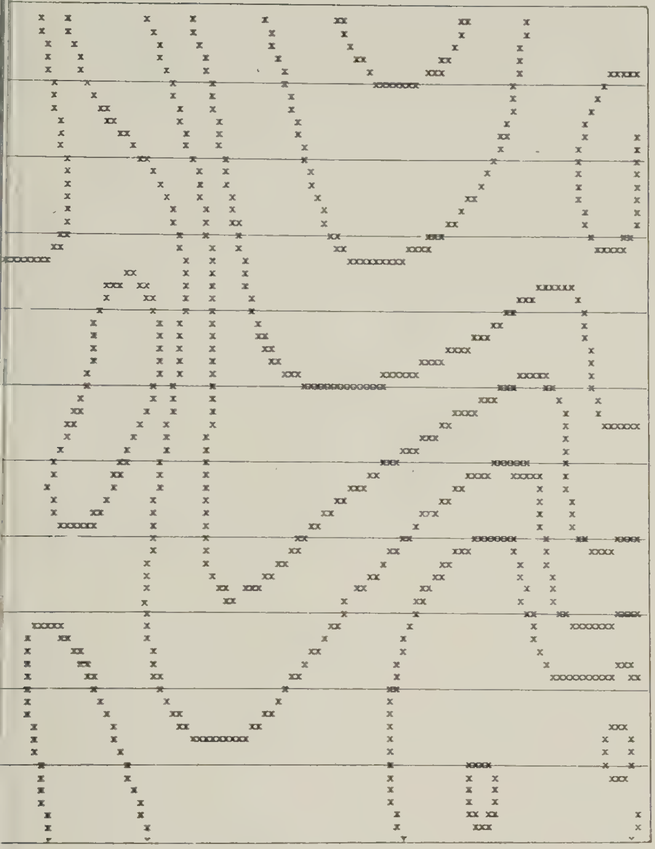


Fig. 1.

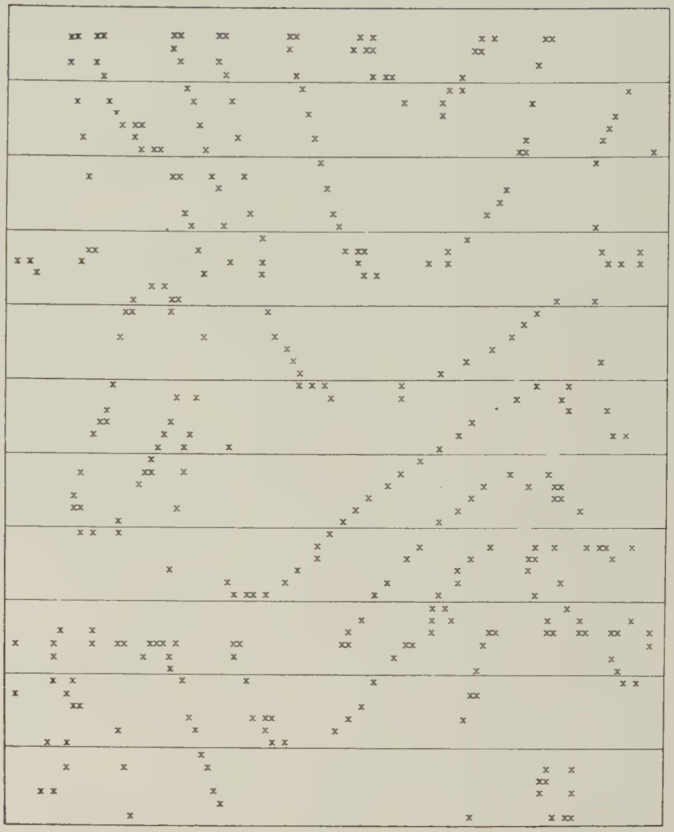


Fig. 2.

TABLE I
PREDICTIVE CODING RESULTS

Map Number	Per cent black elements in picture matrix	Per cent 1's in error matrix	Length of run-length code	Compression coefficient (run-length code / 7000)
1	10.2	4.9	2,394 bits	0.34
2	11.1	5.9	2,798	.40
3	10.8	5.3	2,838	.41
4	10.9	5.9	2,728	.39
5	9.0	4.6	2,240	.32
6	12.6	6.6	2,990	.43
7	11.5	5.7	2,698	.39
8	11.4	5.2	2,585	.37
9	9.5	4.8	2,422	.35
10	14.6	6.5	3,010	.43
Ave.	11.2	5.5	2,670 bits	0.38

We therefore turn to coding methods which could, ther in conjunction with predictive coding or separately, cape the limitations on attainable compression which e inherent in exact coding of pictorial data.

V. Application of Pattern Recognition Techniques

Closely related to the present problem of coding ctorial data is the current work on pattern recognition those interested in psychology and in artificial inlligence. Their work is not restricted to use of the local properties of meaningful pictures, which properties (as mentioned above) do not seem a sufficient basis for a really efficient coding scheme. Meaningful patterns which, in a particular application, would be reacted to in the same way (either because they are not distinguishable by the observer or because, although distinguishable, they provide him with essentially the same information) might as well be considered one pattern and represented by a single code number. Pattern recognition schemes will be useful to us if they provide a means by which, for each

code number, there can be recovered a representative pattern (one of the equivalent patterns having that code number). Our problem is somewhat different from most pattern recognition situations, where each presented pattern has to be recognized as belonging to one of a small set of given categories. We are concerned with the "practically infinite" collection of meaningful patterns, which cannot be given in advance.

In changing our goal from that of assigning short codes to meaningful pictures in a one-to-one manner, where a single change in the matrix representing a picture results in the picture's being considered different and assigned a different code number, we at the same time 1) recognize and compensate for the possibility of noise in the original picture, and 2) reduce the total number of pictures that will be considered distinct, so that average code lengths can be reduced. If, for example, under some scheme of classification, only 2^{2500} meaningful pictures representable on a 500-by-500 matrix are considered different, an average code length of 2500 bits can be aimed for.

Minsky[9,10] feels that a solution of the pattern recognition problem can be obtained without resorting to the use of statistics. He speaks of a program which would isolate each figure in a picture, put the figure into standard form by translation and magnification, analyze the figure for the purpose of recognizing it (or, in our case, coding it), and finally give the geometrical interrelations of the figures which characterize the picture as a whole. The output of the program (for us) would be 1) a set of code numbers, from which each of the figures in the picture could be regenerated, and 2) another code number giving the geometrical relations among the figures.

A first step in the analysis of particular figures can be based on the work of Selfridge[11–13] and Dinneen.[14] Selfridge was interested in finding ways of reducing given patterns to coded versions thereof, picking out the significant features of a pattern in order to classify it. The corresponding computer operations that Dinneen used made possible the removal of small bits of noise and the selection of contour points (edge points of the figures in the picture) and, in particular, the points at which the contour has greatest curvature.

Attneave has given us two schemes which look as if they should be very useful in coding pictorial data efficiently. In the first,[15] he points out a way of taking advantage of the fact that "information is concentrated along contours and is further concentrated at those points on a contour at which its direction changes most rapidly": a good likeness of an object is obtainable by finding the points of high curvature on the boundary of the figure and replacing the curves connecting these points by straight line approximations. We have tried a procedure like this in conjunction with predictive coding (see Section V, A).

A second coding procedure, recommended by Attneave and Arnoult,[16] has not yet been tested by us, but seems even more promising, since it provides curvilinear approximations to given figures. In this procedure, tangents are drawn to a figure at points of low curvature and at corner points. The angles in the resulting polygon are then rounded off, using certain standard arcs to approximate the given curves. Each figure is coded by specifying a starting point, changes in direction in degrees and changes in logarithm of length for the tangent lines forming the polygon, and amount of rounding off for each angle of the polygon.

The recently reported work of Unger[17] on "edge sequences" has proved to be quite useful for our purposes. The edge sequence, which gives essentially the directions (limited to 45°, 90°, $\cdots$, 360°) of the successive line segments which form the boundary of the figure under consideration, can be used as a first "character"[10] for classifying the figure. It can easily be seen that edge sequences will not always be enough to distinguish two figures, but they can be used to help detect similarities among the figures in the picture and the figures in the picture which may fall into one of a *given* (nonexhaustive) set of frequently used categories (*e.g.*, individual letters, numerals, special map symbols). These categories can be given short code designations. A figure which is similar to another in the picture or in the *given* set can be coded by specifying only its location, size, orientation, the number of the figure it is similar to, and the corrections (additions or subtractions) which are necessary to change one figure into the other.

An experiment combining Unger's and Attneave's ideas is described in Section V, B.

Mention should be made, finally, of the work of Uhr,[18] who reported to the 1959 ACM meeting on a program in progress which would have the computer "process forms according to a procedure that recognizes successively higher order relations between successively larger elements, or subwholes, of a form." His program recognizes relative and absolute sizes of lengths, loops, and angles, and can be so used that it makes distinctions only to the rather small degree that is within the abilities of the human observer (5 to 15 just-noticeable-differences along

[9] M. L. Minsky, Lecture given at Mass. Inst. Tech., Cambridge, Mass.; 1958.

[10] M. L. Minsky, "Heuristic Aspects of the Artificial Intelligence Problem," Lincoln Lab., Mass. Inst. Tech., Lexington, Mass.; 1956.

[11] O. G. Selfridge, "Pattern recognition and modern computers," *Proc. WJCC*, Los Angeles, Calif., pp. 82–84; March 1–3, 1955.

[12] O. G. Selfridge, "Pattern recognition and learning," presented at Symp. on Information Theory, London, Eng.; 1955.

[13] O. G. Selfridge, "Pandemonium: A Paradigm for Learning," Lincoln Lab., Mass. Inst. Tech., Lexington, Mass., JA 1140; 1958.

[14] G. P. Dinneen, "Programming pattern recognition," *Proc. WJCC*, Los Angeles, Calif., pp. 94–100; March 1–3, 1955.

[15] F. Attneave, "Some informal aspects of visual perception," *Physchol. Rev.*, vol. 61, pp. 183–193; 1954.

[16] F. Attneave and M. D. Arnoult, "The quantitative study of shape and pattern perception," *Psychol. Bull.*, vol. 53, pp. 452–471; 1956.

[17] S. H. Unger, "Pattern detection and recognition," Proc. IRE, vol. 47, pp. 1737–1752; October, 1959.

[18] L. Uhr, "Machine Perception of Printed and Handwritten Forms by Means of Procedures for Assessing and Recognizing Gestalts," 14th Natl. Meeting of the Assn. for Computing Machinery, Cambridge, Mass.; 1959.

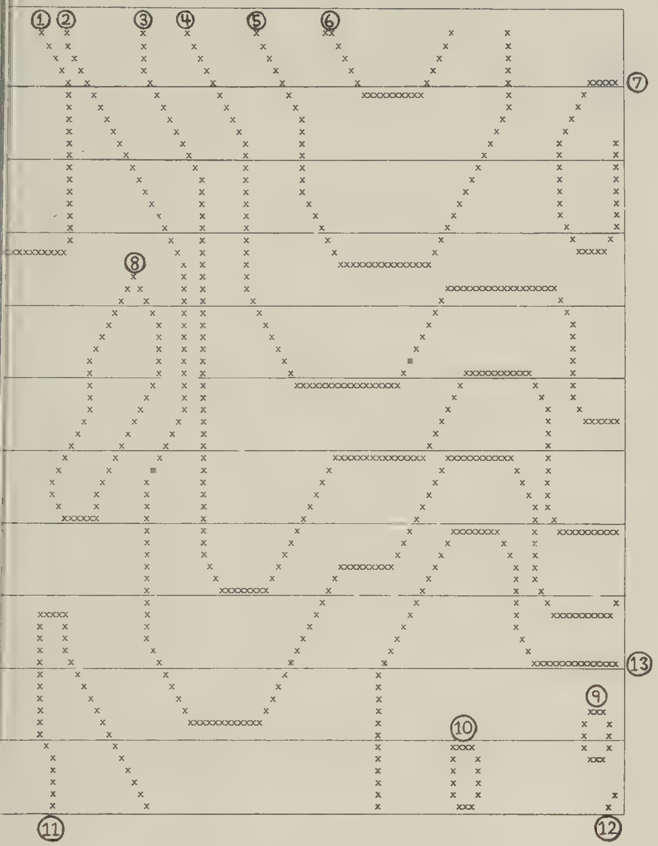


Fig. 3.

Fig. 4.

TABLE II

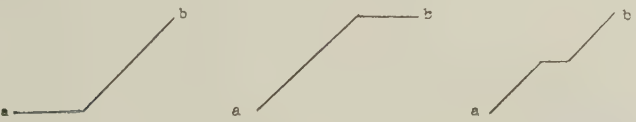
	Per cent 1's in error matrix	Length of run-length code	Compression coefficient $\left(\frac{\text{run-length code}}{7000}\right)$
Map 1	4.9	2,394 bits	0.34
Approximation	1.5	930	.13

ch "dimension": *e.g.*, slope, length, average curvature). his program, or some combination of its operations with ose suggested by Attneave and Arnoult, should provide foundation for approximate coding methods even more ficient than those now to be described.

V. Results of Experiments

. "*Straight-Line*" *Approximations and Predictive Coding*

Using one of the weather maps mentioned in Section I, Selfridge and Dinneen's edging process (in which the adial asymmetry" about each black element is evalu- ed) was used to pick out the points of high curvature on e boundary of each figure on the map. Then, following nly) the spirit of Attneave's suggestion (since straight- e approximation to the curves joining these points ould have become rather broken lines in their matrix presentation), the curve joining each pair of these points high curvature was replaced by an approximation made of a small number of horizontal, vertical, and diagonal ith slope ± 45°) line segments. To connect the points and *b*, for example, broken lines of the following types ere used.



none of these broken lines gave a close approximation to e original curve joining the two points (such a failure uld have been detected mechanically), an intermediate point of the original curve was selected and joined to each of the originally chosen points by one of the three kinds of broken lines. The resulting map appears in Fig. 3.

The predictive coding program (with twelve-element neighborhoods) was then applied quite successfully to the standardized map thus obtained. There is a striking difference between the original error matrix, which looked a great deal like the map itself, and the error matrix resulting from predictive coding applied to the standardized map (compare the print-outs in Figs. 2 and 4). Whereas run-length coding of the original error matrix had resulted in a code of length 2,394, from which the original map could be recovered, run-length coding of the error matrix corresponding to the standardized version of the map resulted in a code of length 930, from which only the standardized approximation to the original map can be recovered (Table II).

B. "*Straight-Line*" *Approximations and Edge Sequences*

The standardized map obtained in the previous experiment could be coded, not by predictive coding, but by a more complex code specifying the location and

description of each of the figures in the picture. Each figure (unless it could be recognized as belonging to one of the *given* categories, *e.g.*, circle or "*A*") would be coded by specifying the directions and lengths of the successive line segments which formed the figure. If a figure were "similar" to (*i.e.*, had roughly the same shape as) another in the map or in a given category, that fact would be noted and used to predict both its edge sequence and the lengths of the edges. The fact that one figure was known to be similar to another would make a great deal of the direction and length information redundant and therefore allow compression into a shorter code.

A rough calculation of the code lengths necessary to reproduce essentially the standardized map of the previous section yielded the following estimate (which should be looked upon as an upper bound):

1) *To Specify Starting Points (or Centers, in the Case of Circles) for the Thirteen Figures in the Standardized Map of Fig.* 3: Since most of the 13 points fall on the edges of the map, we could code them very efficiently by run-length coding if we ordered the 7000 elements of the matrix in the following way:

1	2	3	–	–	–	100
336	337	–	–	–	–	101
–	–	–	–	–	.	–
–	–	–	–	–	–	–
–	–	–	–	–	7000	–
268	–	–	–	–	170	169.

About 110 bits would be required.

2) *To Designate for Each Figure*: a) to which of the 13 figures it is similar (for this computation figures were considered similar if their *edge sequences* were similar; we allow, in particular, that a figure be similar only to itself) or b) to which of the *given* categories it belongs ("circle" being the only category used here). About $13 \times 4 = 52$ bits would be required ($\log_2 (13+1) \doteq 4$).

3) *To Specify the Size (Radius) of Each of the Two "Circles"*: about $2 \times 4 = 8$ bits, allowing the radius to have any one of (say) 16 possible lengths. (Note the great reduction possible when the figure falls into a *given* category.)

4) *To Specify (Direction-of-) Edge Sequences for the Six "Similar" Figures (by Comparison of Edge Sequences, Fig. 5 is a similar to Fig.* 6, 4 *to* 5, 3 *to* 4, 2 *to* 3, *and* 13 *to* 2): About 60 bits for run-length coding of the error sequence resulting from use of predictive coding on the edge sequences, where predictions ($\pm$ 45° turns) would be made on the basis of the edge sequence of the pattern to which the pattern under consideration was similar (when this could not be used, predictions would be made on the basis of previous curvature).

5) *To Specify (Length-of-) Edge Sequences for the Six "Similar" Figures*: About 3 bits/edge: about 165 bits (length of the corresponding segment, if there was one in the similar pattern, being used to get approximate length).

6) *To Specify (Direction-of-) Edge Sequences for the Five "Nonsimilar" Figures*: About 65 bits for the error sequence resulting from use of predictive coding on the edge sequence, predictions ($\pm$ 45° turns) being made (not very successfully) on the basis of continuation of the previous curvature.

7) *To Specify (Length-of-) Edge Sequences for the Five "Nonsimilar" Figures*: About 4 bits/edge: about 125 bits.

The code length would be, therefore, approximately $110 + 52 + 8 + 60 + 165 + 65 + 125 = 585$ bits, which represents a compression coefficient of $585/7000 \doteq 0.08$.

A computer coding pictorial data in the manner just outlined could even use the following operational definition of similarity. Fig. k is similar to Fig. l (or to the figures in category l) if this designation of Fig. k in 2) yields a shorter code for that figure in 4) and 5) than would be obtained for that figure in 6) and 7).

VI. Conclusion

After discussion of an experiment in which predictive coding was performed by a computer, we moved away from the ideal of exact coding and reproduction of pictures in order to achieve more compression while still preserving the significant features of pictures, maps, etc. Using pattern recognition techniques, with or without subsequently resorting to predictive coding, we have taken account of large-scale features of pictures, avoiding sole reliance on local operations. The operations considered enable a computer to categorize pictures efficiently without its having been given a complete set of categories.

Reductions in the compression coefficient (from 0.34 to 0.13 or 0.08) point toward still greater savings to be realized through the use of pattern recognition methods when realistic high-resolution data is processed, since there it is clearly the general pattern that is important, and not the colors of particular matrix elements.

VII. Acknowledgment

The author is indebted to and here expresses his thanks to R. Gold, under whose direction much of the work here reported on predictive coding was done; to R. Wernikoff; and to R. Schwartz, who wrote the computer programs mentioned above and encouraged the writing of this report.

On Singular and Nonsingular Optimum (Bayes) Tests for the Detection of Normal Stochastic Signals in Normal Noise*

DAVID MIDDLETON, FELLOW, IRE†

ummary—The necessary and sufficient ($n.$ and $s.$) conditions the nonsingularity, *i.e.*, regularity, and for the singularity of mum tests for the presence of one Gaussian process vs another a finite sample are established, for both nonstationary and ionary processes, including those with nonrational spectra. he stationary cases, the condition may be expressed alternatively erms of an integral of suitable spectral ratios when the random cesses possess rational spectra and for certain classes of non-onal spectra as well. Equivalently, for rational spectra the and $s.$ condition for nonsingularity is that the spectral ratio roach unity as frequency becomes infinite and that the spectral o be finite and nonzero for all frequencies, while for singularity $n.$ and $s.$ condition requires that this ratio differ from unity he limit or if unity in the limit, that this ratio vanish or be un-nded at some one (or more) finite frequencies. Some of the lications of these results in applications to signal detection are sidered, and a method of solution of an associated class of egral equations, of the type

$$\int^{T+} L(\tau, u)K(|u - t|)\, du = G(t, \tau), \quad 0- < t, \tau < T+$$

ere K is a rational kernel and G is suitably specified, is briefly lined. Specific results in the case of RC and LRC noise kernels, h G correspondingly the difference of two (different) RC or LRC ariance functions, are also given.

I. Introduction

THE problem of detecting optimally the presence of a Gaussian signal in normal noise has been considered by the author[1,2] and a number of other estigators in recent years.[3] The results are generally nsingular Bayes (*i.e.*, minimum average risk) tests, t is, optimum statistical tests which yield nonzero and nunity probabilities of error, based on finite samples. wever, as Slepian[4] has more recently pointed out, imum singular tests, whereby perfect detection, bability 1, is possible with arbitrarily small samples, y arise in certain instances, where in effect the math-atical model is poorly chosen to represent the actual physical circumstances.[5–7] In the case of two different stationary Gaussian processes $N_0(t)$, $N_1(t)$, with zero means and spectral intensity densities $\mathcal{W}_0(f)$, $\mathcal{W}_1(f)$, respectively, Slepian gives some sufficient conditions for the singularity of such tests. The exceptional cases, about which his theorem makes no statement, occur when $\lim_{f\to\infty} \mathcal{W}_1(f)/\mathcal{W}_0(f) = 1$. The main purpose of the present paper is to state and to outline briefly the derivation of the necessary and sufficient ($n.$ and $s.$) condition for the existence of the optimum nonsingular (Bayes) test of one Gaussian process $N_1(t)$ vs another $N_0(t)$ for 1) general, nonstationary processes and 2) various classes of stationary processes, where both N_1 and N_0 have zero means.[8] From this we obtain also the necessary and sufficient conditions for 1) and 2), that the optimum test N_1 vs N_0 be singular, accounting in both instances as well for the important exceptional case in Slepian's theorem.[4] Several comments on the implications of these results for physical applications complete our discussion.

II. Principal Results

In this section we summarize the principal results formally as a series of theorems and in Sections III-VI provide the details of the proofs. We begin by postulating that the random processes in question, N_0 and N_1, are

1) *normal, with zero means;*
2) *possess covariance functions K_0, K_1 that are positive definite, symmetrical, continuous and quadratically integrable on an arbitrary observation interval* $(0-, T+)$.

Then, we have the following theorems, subject to various additional assumptions on the processes N_0, N_1 such as stationarity, nonstationarity, spectral rationality and nonrationality. We consider first the general case:

* Received by the PGIT, June 29, 1960; revised manuscript eived, September 19, 1960. This paper is based on Group Rept. -0001, of the same title, Lincoln Lab., Mass. Inst. Tech., ington, Mass.; June, 1960.

† Consultant, Lincoln Lab., M.I.T., Lexington, Mass., operated h support from the U. S. Army, Navy, and Air Force.

D. Middleton, "On the detection of stochastic signals in additive mal noise, pt. I," IRE Trans. on Information Theory, vol. 3, pp. 86–121; June, 1957, 256, 257; December, 1957.

D. Middleton, "An Introduction to Statistical Communication ory," Internatl. Ser. in Pure and Appl. Phys., McGraw-Hill k Co., Inc., New York, N. Y., Sect. 20.4–7; 1960.

Middleton, *op. cit.*, Reference 1, Bibliography, especially [2]–[4],

D. Slepian, "Some comments on the detection of Gaussian als in Gaussian noise," IRE Trans. on Information Theory, IT-4, pp. 65–68; June, 1958.

[5] U. Grenander, "Stochastic processes and statistical inference," *Arkiv. Math.* vol. 1, p. 195; 1950. For a first example involving the case of normal noise vs normal noise, *cf.* Sect. 4. For an example in sequential detection, see[6].

[6] J. J. Bussgang and D. Middleton, "Optimum sequential detection of signals in noise," IRE Trans. on Information Theory, vol. IT-1, p. 5; December, 1955.

[7] I. J. Good, "Effective sampling rates for signal detection: or can the Gaussian model be salvaged?" *Information and Control*, vol. 3, p. 116; June, 1960. For conditions governing the avoidance of singular situations in the information capacity of certain types of channels, in reply to Good's work, see P. Swerling, "Pardoxes related to the rate of transmission of information," *Information and Control*, vol. 3, p. 351; December, 1960.

[8] For nonvanishing means, our subsequent argument is simply modified without any essential changes in procedure.

Theorem 1(a): The necessary and sufficient condition that the optimum (Bayes) test of N_1 vs N_0 on $(0-, T+)$ be regular, i.e., nonsingular, is that

$$-\infty < \sum_1^\infty \lambda_j^{(ab)} = \int_{0-}^{T+} L_{ab}(t, t)\, dt < \infty; \qquad \begin{matrix} a = 1, & b = 0 \\ a = 0, & b = 1 \end{matrix} \tag{1}$$

where the functions $L_{ab}(t, u)$ are the solutions of the inhomogeneous integral equations

$$\int_{0-}^{T+} L_{ab}(t, u)K_b(u, \tau)\, du = K_1(t, \tau) - K_0(t, \tau), \qquad 0 \leq t, \tau \leq T. \tag{2}$$

Here $\lambda_j^{(ab)}$ are the eigenvalues of the associated homogeneous integral equations

$$\int_{0-}^{T+} L_{ab}(t, u)\phi_j^{(ab)}(u)\, du = \lambda_j^{(ab)}\phi_j^{(ab)}(t), \quad 0 \leq t \leq T. \tag{3}$$

Theorem 1(b): The necessary and sufficient condition that the optimum (Bayes) test of N_1 vs N_0 on $(0-, T+)$ be singular, i.e., yield perfect detection (probability 1) for all finite $T > \epsilon \geq 0$, is that

$$\sum_1^\infty \lambda_j^{(ab)} = \int_{0-}^{T+} L_{ab}(t, t)\, dt \quad \textit{diverges}. \tag{4}$$

We remark that the eigenvalues of (3) are not necessarily real, since $L_{ab}(t, \tau) \neq L_{ab}(\tau, t)$ in general. We also observe that Theorems 1(a) and 1(b) apply for stationary as well as nonstationary normal processes, with either rational or nonrational spectra.

In the stationary situations, we have the following alternative theorems, equivalent to Theorems 1(a) and 1(b), but now expressed in terms of the respective spectral intensity densities of N_0, N_1. We consider first the cases involving rational spectra, *i.e.*, spectra of the processes obtained by passing white (normal) noise through an invariant, stable, lumped-constant network:

Theorem 2(a): If the normal processes N_0, N_1 are stationary with rational spectra $\mathcal{W}_0$, $\mathcal{W}_1$, the necessary and sufficient condition that the optimum (Bayes) test of N_1 vs N_0 on $(0-, T+)$ be regular i.e., nonsingular, is that

$$\left| 2T \int_0^\infty \left[\frac{\mathcal{W}_1(f)}{\mathcal{W}_0(f)} - \frac{\mathcal{W}_0(f)}{\mathcal{W}_1(f)}\right] df \right| < \infty, \qquad \text{all } 0 < T < \infty, \tag{5a}$$

or equivalently, that

$$\lim_{f\to\infty} \frac{\mathcal{W}_1(f)}{\mathcal{W}_0(f)} = \lim_{f\to\infty} \frac{\mathcal{W}_0(f)}{\mathcal{W}_1(f)} = 1, \quad \text{all } 0 < T < \infty, \tag{5b}$$

and that

$$0 < \frac{\mathcal{W}_1(f)}{\mathcal{W}_0(f)}, \quad \frac{\mathcal{W}_0(f)}{\mathcal{W}_1(f)} < \infty, \quad \text{all } 0 \leq f < \infty.$$

The equivalence of (5a) and (5b) follows at once from the fact that for all rational spectra under the latter conditions there can be no spectral "holes" in $\mathcal{W}_0$, $\mathcal{W}_1$ over any finite frequency interval, alternatively reflected in the fact that the covariance functions K_a, K_1 are positive definite (as opposed to positive *semi*-definite).

Theorem 2(b): The necessary and sufficient condition for the singularity of the optimum (Bayes) test of N_1 vs N_0 when N_1, N_0 possess rational spectra $\mathcal{W}_1$, $\mathcal{W}_0$ respectively, is that for all $0 < T < \infty$

$$2T \int_0^\infty \left[\frac{\mathcal{W}_1(f)}{\mathcal{W}_0(f)} - \frac{\mathcal{W}_0(f)}{\mathcal{W}_1(f)}\right] df \quad \textit{diverges}. \tag{6a}$$

This is equivalent to

$$\lim_{f\to\infty} \frac{\mathcal{W}_1(f)}{\mathcal{W}_0(f)} \neq 1, \quad \textit{and/or} \quad \frac{\mathcal{W}_1(f)}{\mathcal{W}_0(f)} = 0 \tag{6b}$$

for some $f(0 \leq f < \infty)$.

(Note that it is possible to have $\lim_{f\to\infty} \mathcal{W}_1 f/\mathcal{W}_0(f) = 1$ and still have singularity, if $\mathcal{W}_1/\mathcal{W}_0$ vanishes or is unbounded in $0 \leq f < \infty$.)

However, in many physical situations the noise mechanism is in some sense represented by distributed sources so that the resulting spectrum is nonrational; clutter scatter multipath, noise generated in traveling-wave tubes are important examples. For some of these nonrational cases we may state the following:

Theorem 3: In the special cases for which the nonrational spectra may be regarded as suitable limiting forms of rational spectra, the necessary and sufficient condition[9] for regularity or singularity of the optimum (Bayes) test of N_1 vs N_0 is now that (5a) *or* (6a) *apply, respectively.*

For the special class of nonrational spectra characterized by bandlimiting, where N_0, N_1 are bandlimited to the same (or different) spectral intervals, we have the following theorem:

Theorem 4: A sufficient condition that the optimum (Bayes) test of N_1 vs N_0 on $(0-, T+)$ be singular is that N_1, or N_0, or both N_1 and N_0, be bandlimited to the same or different spectral regions.

Note that bandlimiting is a sufficient, but not a necessary condition for singularity, which may also occur for rational spectra, of Theorem 2(b) above. Theorem 4 is essentially Slepian's result,[4] as is (6b) from the viewpoint of sufficiency. Finally, we observe that for both the singular and regular cases in all instances the conditions (1), (4), (5a)–(6b) are qualitatively independent of interval length T, as long as this interval length is finite. Some of the implications of these results in applications are discussed in Section VII.

[9] The nonrational cases are fully covered by Theorems 1(a) and 1(b). We were not, however, able here to demonstrate necessity and sufficiency for the spectral form of these theorems in the general nonrational situations, although on physical grounds we strongly suspect that (5a) and (6a) represent both necessary and sufficient conditions, respectively, for regularity and singularity in the case of spectra which are not bandlimited.

II. The General Situation; Proof of Theorems 1(a) and 1(b)

We consider first the general situation described at the [b]inning of Section II above, where no specific assump[tio]ns of stationarity, spectral rationality, etc. are neces[sar]ily made. To establish the results embodied in Theorems [1(a)] and 1(b) of (1), (4), we start with the discrete case [of] n sampled values $\mathbf{V} = [V_1, V_2, \cdots, V_n]$ of the received [pro]cess $V(t)$ on $(0, T)$. For the gauss processes postulated [her]e, the optimum detector structure for the Bayes test [of] N_1 vs N_0 is the generalized likelihood ratio.[10]

$$\log \Lambda_n = \log \mu_{10} - \tfrac{1}{2} \log \det \mathbf{K}_1\mathbf{K}_0^{-1} - \tfrac{1}{2}\tilde{\mathbf{V}}(\mathbf{K}_1^{-1} - \mathbf{K}_0^{-1})\mathbf{V}, \qquad (7)$$

[wh]ere $\mu_{10} = p_1/p_0$ is the ratio of *a priori* probabilities that [th]e sample $\mathbf{V}$ represents N_1 or N_0, with $p_0 + p_1 = 1$. Here [$\mathbf{K}_1$], $\mathbf{K}_0$ are the $n \times n$ covariance matrices of N_1 and N_0 [at] times $t = t_1, \cdots, t_n$ in $(0, T)$, with the nonsingular [in]verses $\mathbf{K}_1^{-1}, \mathbf{K}_0^{-1}$.

For the regularity of the test we must show with con[tin]uous sampling on $(0-, T+)$ that the logarithm of the [lik]elihood ratio *functional* $\log \Lambda_T = \lim_{n\to\infty} \log \Lambda_n$ is [bo]unded and approaches a different unique value with [re]spect to each hypothesis $H_0 \in N_0$, $H_1 \in N_1$, *and* that the [co]nditional error probabilities $\beta_0^{(1)}, \beta_1^{(0)}$ of deciding N_0 [wh]en N_1 is actually present, and vice versa, are neither [ze]ro not unity, *i.e.*, that in the limit $n \to \infty$, $0 < \beta_0^{(1)}$, [$\beta_1^{(0)}$] < 1 for all $0 < T < \infty$. The necessary and sufficient [co]ndition for this is established from the demonstration [its]elf, as will be noted presently.

Accordingly, we must show that

$$-\infty < \lim_{n\to\infty} \mathbf{E}_{\mathbf{V}|H_a}\{\tilde{\mathbf{V}}(\mathbf{K}_1^{-1} - \mathbf{K}_0^{-1})\mathbf{V}\} < \infty, \qquad a = 0, 1, \qquad (8a)$$

[an]d that

$$|B_T| \equiv |\lim_{n\to\infty} \{\log \mu_{10} - \tfrac{1}{2} \log \det \mathbf{K}_1\mathbf{K}_0^{-1}\}| < \infty. \qquad (8b)$$

[N]ext, we observe that $\log \Lambda_T$, *cf.* (7) in the limit, may [co]ntain either a positive definite, or a negative definite [qu]adratic functional of $V(t)$, or a quadratic functional of [V](t) that is neither positive nor negative definite. The [di]stribution densities (d.d.'s) $P_1(x), P_0(x)$ of the functional [x] $\equiv \log \Lambda_T$ with respect to H_1, H_0 accordingly vanish for [all] $x < x_1, x < x_0$, or for all $x > x_1, x > x_0$, respectively [as] x contains a positive or negative definite quadratic [fu]nctional, where x_0, x_1 are appropriate limits for the [ra]nge of values of the random variable $x = \log \Lambda_T$ in the [tw]o hypothesis cases H_0, H_1. However, when the quadratic [fu]nctional in $\log \Lambda_T$ is neither positive nor negative [de]finite, the limits x_0, x_1 do not exist, and $P_1(x), P_0(x)$ [ma]y be expected to be different from zero for regions of x [an]ywhere in $(-\infty\ x < \infty)$. In fact, in order to complete [th]e demonstration by showing that $0 < \beta_0^{(1)}, \beta_1^{(0)} < 1$ in the limit $n \to \infty$, we must not only verify that the distribution densities $P_1(x), P_0(x)$ have no singularities, *i.e.*, the associated cumulative distributions possess no "mass points" anywhere, but we must also show that $P_1(x), P_0(x)$ are everywhere continuous, for $x > x_1, x > x_0$ in the case of a positive definite quadratic functional in $V(t)$, and that in this case $\mathbf{E}_{H_0}\{x\} \equiv \bar{x}_0 > x_0$, $\mathbf{E}_{H_1}\{x\} \equiv \bar{x}_1 > x_1$, necessarily, with no finite regions in $(x_{0,1} < x < \infty)$ where P_0, P_1 are zero. With negative definite quadratic functionals, these regions are reversed, and when $\log \Lambda_T$ is neither positive nor negative definite, the requirement is that $P_1(x), P_0(x)$ are continuous, all $x\ (-\infty < x < \infty)$. Then, in each instance if $\bar{x}_0, \bar{x}_1$ are finite, $\beta_0^{(1)}, \beta_1^{(0)}$ can be neither zero nor unity, and the optimum test ($\log \Lambda_T \geq \log \mathcal{K}$ for H_1, $\log \Lambda_T < \log \mathcal{K}$ for H_0, where $\mathcal{K}$ is some finite threshold) is accordingly regular.

We begin with the quadratic form $\Phi_n \equiv \tilde{\mathbf{V}}(\mathbf{K}_1^{-1} - \mathbf{K}_0^{-1})\mathbf{V}$ and write it alternatively

$$\Phi_n = \tilde{\mathbf{V}}\mathbf{K}_1^{-1}(\mathbf{I} - \mathbf{K}_1\mathbf{K}_0^{-1})\mathbf{V} = -\tilde{\mathbf{V}}\mathbf{K}_1^{-1}\mathbf{H}_{10}\mathbf{V} \qquad (9a)$$

or

$$= \tilde{\mathbf{V}}\mathbf{K}_0^{-1}(\mathbf{K}_0\mathbf{K}_1^{-1} - \mathbf{I})\mathbf{V} = -\tilde{\mathbf{V}}\mathbf{K}_0^{-1}\mathbf{H}_{01}\mathbf{V}, \qquad (9b)$$

where

$$\mathbf{H}_{10} \equiv \mathbf{K}_1\mathbf{K}_0^{-1} - \mathbf{I}; \qquad \mathbf{H}_{01} \equiv \mathbf{I} - \mathbf{K}_0\mathbf{K}_1^{-1} \qquad (9c)$$

define $\mathbf{H}_{10}, \mathbf{H}_{01}$. From (9c) we get directly the basic relations

$$\mathbf{H}_{ab}\mathbf{K}_b = \mathbf{K}_1 - \mathbf{K}_0; \quad a = 1, b = 0; a = 0, b = 1. \qquad (10)$$

Now, consider the expectations

$$\Phi_n^{(a)} = -\sum_{ij} \mathbf{E}_{\mathbf{V}|H_a}\{V_iV_j\}(\mathbf{K}_a^{-1}\mathbf{H}_{ab})_{ij} = -\text{trace } \mathbf{H}_{ab}. \qquad (11)$$

Next, let $H_{ab} \equiv L_{ab}\Delta t$, $\Delta t = T/n$ and pass to the limit[11] $(n \to \infty)$, obtaining for (11)

$$\Phi_T^{(a)} = -\int_{0-}^{T+} L_{ab}(t, t)\, dt \qquad (12)$$

where the L_{ab} are now determined from the pair of basic integral equations, obtained from the limit of (10), *viz.* (2) above.

The bias B_T, Eq. (8b), similarly becomes

$$B_T = \log \mu_{10} - \tfrac{1}{2} \lim_{n\to\infty} \log \det (\mathbf{I} + \mathbf{K}_1\mathbf{K}_0^{-1} - \mathbf{I})$$

$$= \log \mu_{10} - \tfrac{1}{2} \lim_{n\to\infty} \log \det (\mathbf{I} + \mathbf{H}_{10})$$

$$= \log \mu_{10} - \tfrac{1}{2}\, \mathfrak{D}_{10}(1) = \log \mu_{10} + \tfrac{1}{2}\, \mathfrak{D}_{01}(-1), \qquad (13)$$

where $\mathfrak{D}_{10}(1)$, $\mathfrak{D}_{01}(-1)$ are the Fredholm determinants

[10] Middleton, *op. cit.*, Reference 2, (19.20).

[11] This could also be expressed as a Stieltjes integral. Necessary and sufficient conditions for the existence of solutions L_{ab} of (2) are that K_1, K_0 obey the conditions assumed above for positive definiteness, etc. For if we fix t, for the moment, in (2) we see that the resulting integral equation is a special class of a more general inhomogeneous type treated earlier, *cf.* Sect. 19.4–2 of Middleton, *op. cit.*, Reference 2, the discussion therein, and references.

$\prod_{j=1}^{\infty}(1+\lambda_j^{(10)})$, $\prod_{j=1}^{\infty}(1-\lambda_j^{(01)})$, and $\lambda_j^{(ab)}$ are the eigenvalues of the associated homogeneous equation (3).[12] Note that the $\Phi_T^{(a)}$ are, in effect, the first iterated kernels of (3), and so we have also

$$\Phi_T^{(a)} = -\sum_{j}^{\infty} \lambda_j^{(ab)}, \qquad \begin{matrix} a = 1, b = 0 \\ a = 0, b = 1 \end{matrix}. \tag{14}$$

Consequently, for $\Phi_T^{(a)}$ and B_T to exist, it is sufficient that $\sum_{j=1}^{\infty} \lambda_j^{(ab)}$ be finite, *cf.* (1), since the convergence of the Fredholm determinent is insured by the convergence of the series (14).[13]

To complete the proof, we need next to consider the characteristic functions (c.f.'s) of $P_0(x)$ and $P_1(x)$. We start here with the c.f.'s for log Λ_n, which are readily found to be[14]

$$F_a(i\xi)_n = \frac{e^{i\xi B_n}}{\{\det(\mathbf{I} - i\xi\mathbf{H}_{ab})\}^{1/2}} \tag{15}$$

which for $n \to \infty$ become

$$F_a(i\xi)_T = e^{i\xi B_T}\, \mathfrak{D}_{ab}(-i\xi)^{-1/2} \tag{16}$$

for P_a, $a = 0, 1$, where

$$\mathfrak{D}_{ab}(-i\xi)^{-1/2} = \prod_{j=1}^{\infty} (1 - i\xi\lambda_j^{(ab)})^{1/2}.$$

But, by a theorem of Gnedenko and Kolmogoroff,[15] the P_a cannot be the distribution densities (d.d.'s) of discrete (or lattice) distributions, *i.e.*, the distributions associated with the densities P_a cannot have "mass points" since clearly $F_a(i\xi)_T \neq 1$ for all $\xi \neq 0$. Moreover, there cannot be more than one region where $P_a = 0$, when $-\Phi_T$ is positive or negative definite, and this region occurs for $x < x_0, x_1$ or $x > x_0, x_1$, respectively. Also, when Φ_T is neither positive nor negative definite, there is no region for which P_a vanishes, $-\infty < x < \infty$. Consequently, the distribution densities P_a, $a = 0, 1$, are continuous and bounded, and, therefore, $0 < \beta_0^{(1)}, \beta_1^{(0)} < 1$ as required for nonsingularity.

Further, it is easily established that all moments of these d.d.'s exist under H_0, H_1. For, differentiating the c.f.'s (16) with the help of the author's expansion[16] and setting $\xi = 0$ in the result, yield[17]

$$\bar{x}_a = B_T + \tfrac{1}{2}\sum_{j=1}^{\infty} \lambda_j^{(ab)};$$

$$\overline{x_a^2} - \bar{x}_a^2 = \tfrac{1}{2}\sum_{j=1}^{\infty} [\lambda_j^{(ab)}]^2, \quad \text{etc.} \tag{17}$$

By comparing the higher-order semi-invariants with the corresponding moments, it is readily seen that $\overline{x_a^m} = 0(\bar{x}_a^m)$, and if $|\bar{x}_a| < \infty$, *i.e.*, if once more $\sum_j \lambda_j^{(ab)}$ is convergent, all moments accordingly exist.

Now in fact, we can establish the necessity and sufficiency of the condition $|\sum_1^{\infty} \lambda_j^{(ab)}|$ by considering the series $\sum_j \lambda_j^{\nu}$, [$\lambda_j = \lambda_j^{(ab)}$ for brevity] where $0 < \nu$. For suppose $\nu \geq 1$; then since the mth semi-invariants are $\sum_j \lambda_j^m$, all semi-invariants ($m \geq 1$) are defined. This is clearly a necessary condition, since $\sum_j \lambda_j$ defines $\bar{x}_a$. Also, if $0 < \nu < 1$ and $\sum_j \lambda_j^{\nu}$ is bounded, we have clearly a sufficient condition, and one in fact that is too strict. But for $\nu = 1$ both sufficiency and necessity meet[18]: the existence of $\bar{x}_a$ is established, and all moments, as well as all semi-invariants, are defined, also insuring the regularity of the test.

The proof of singularity [Theorem 1(b)] now follows immediately: the necessary and sufficient condition for singularity is simply that

$$\left|\sum_1^{\infty} \lambda_j\right| \to \infty$$

with respect to hypotheses H_0, H_1, by obvious substitutions of divergence for convergence at appropriate places in the preceding demonstration [Theorem 1(a)]. Finally we remark that these results apply for both stationary and nonstationary processes, as long as the fundamental assumptions (1) and (2), Section II, are obeyed.

IV. Alternative Results for Stationary Processes with Rational Spectra; Proof of Theorems 2(a) and 2(b)

When N_0, N_1 are stationary, as well as normal, and possess rational spectra $\mathcal{W}_0$, $\mathcal{W}_1$, alternative forms of Theorems 1(a) and 1(b) maybe obtained in terms of these spectral densities, *cf.* (5a)–(6b). We begin here with the proof of regularity, Theorem 2(a), and then establish the necessary and sufficient conditions for singularity, Theorem 2(b).

To derive necessary equivalent and sufficient conditions (5a), (6), we start with integral equations (2), interchange t and τ therein for convenience, and write (2) as

$$\int_{0-}^{T+} L_{ab}(\tau, u) K_b(|u - t|)\, du = K_1(|t - \tau|)$$

$$- K_0(|t - \tau|) \equiv G(|t - \tau|), \; 0- < t, \tau < T+, \tag{18}$$

where we have used the stationarity property $K_b(t, u) = K_b(|t - u|)$, $b = 0, 1$.

Our next step is to regard τ for the moment as a parameter and observe that (18) is an inhomogeneous Fredholm integral equation (of the first kind) which can be solved

[12] Middleton, *op. cit.*, References 2, (17.5) *et seq.*

[13] *Ibid.*, pp. 725, 726, 730.

[14] *Ibid.*, Sect. 20.4–7 [for example we may follow the procedure indicated in (2) of this Section].

[15] B. V. Gnedenko and A. N. Kolmogoroff, "Limit Distributions of Sums of Independant Random Variables," Addison-Wesley Co., Reading, Mass., p. 5, Theorem 5; 1954.

[16] Middleton, *op. cit.*, Reference 2, (17.19).

[17] *Ibid.*, Sect. 17.2.

[18] In certain cases as J. Feldman ("Equivalence and perpendicularity of Gaussian processes," *Pacific J. Math.*, vol. 8, p. 699; 1958) has shown, necessity and sufficiency are obtained when $\sum_j \lambda_j^2$ is bounded, while $\sum_j \lambda_j$ can diverge. Here, the bias B_T and data operator Φ_T both diverge in such a way that the test yields finite nonzero and nonunity probabilities of error. However, for practical applications this situation is clearly unrealizable, as it yields an unspecifiable system structure (embodied in B_T and Φ_T) so that the convergence, or divergence, of $\sum_j \lambda_j$ remains the significant condition.

rational kernels, in the manner of the author.[19] The tional kernels K_0, K_1 by the Wiener Khintchine theorem ve the corresponding rational spectra.[20]

$$W_0(f) = C_0 \prod_{n=0}^{M_0} (c_n^{(0)^2} + \omega^2) \Big/ \prod_{n=1}^{N_0} (b_n^{(0)^2} + \omega^2)$$

$$= 4 \sum_{n=1}^{N_0} \frac{a_n^{(0)} b_n^{(0)}}{b_n^{(0)^2} + \omega^2}, \qquad \omega = 2\pi f, \qquad (19a)$$

d

$$W_1(f) = C_1 \prod_{n=0}^{M_1} (c_n^{(1)^2} + \omega^2) \Big/ \prod_{n=1}^{N_1} (b_n^{(1)} + \omega^2)$$

$$= 4 \sum_{n=1}^{N_1} \frac{a_n^{(1)} b_n^{(1)}}{b_n^{(1)^2} + \omega^2}, \qquad (19b)$$

ith C_0, $C_1 > 0$, $Re\ (b_n^{(0)}, b_n^{(1)}) > 0$, $N_{0,1} \geq M_{0,1} + 1$, here for $n = 0$, $c_n^2 + \omega^2$ is replaced by unity.[21] In addition, (t, τ) is required to be differentiable on $(0- < t, \tau < T+)$ the appropriate order, and for the moment we assume l spectral poles and zeros to be simple and distinct, a striction easily removed at later stages by the observation that the cases of multiple poles may be treated by suitable limiting process, *e.g.*, $b_n \to b_{n+2}$, etc.[22] With the bove modifications and assumptions the general procedure for solving (18) is described in detail elsewhere,[23] ith examples.[24] Here, however, we shall first somewhat odify the approach referred to above, in order to reveal irectly the spectral forms (5), and then return to this riginal approach for explicit evaluations of $L_{ab}\ (\tau, t)$ $_{ab}(t, \tau)$ etc.

Since the kernel K_0 in (18) is continuous everywhere, t us set $G^{(+)} \equiv G(T, \tau)$, $G^{(-)}(\tau) \equiv G(0, \tau)$, and following e steps given elsewhere,[25] let us rewrite (18) for $a = 1$, $= 0$ as

$$\int_{-\infty}^{\infty} K_0(|\,t - u\,|)L_{10}(\tau, u)_T\, du = G(t, \tau)$$

$$+ \begin{cases} G^{(+)}(\tau) \exp\,[-c^{(+)}(t - T)] - G(t, \tau) + \mathfrak{X}^{(+)}(t, \tau), & t > T \\ 0, & 0 \leq t \leq T \\ G^{(-)}(\tau) \exp\,[c^{(-)}t] - G(t, \tau) + \mathfrak{X}^{(-)}(t, \tau), & t < 0 \end{cases} \qquad (20)$$

where $L_{10}(\tau, t)_T$ vanishes outside $(0- < t < T+)$ and is equal to $L_{10}\ (\tau, t)$ inside, and where $\mathfrak{X}^{(+)}$ is zero for $t \leq T$, and $\mathfrak{X}^{(-)}$ vanishes for $t \geq 0$. The constants $c^{(\pm)}$ are to be chosen presently [*cf.* (23)] and the exponential factors, along with the known, convergent behavior at $\pm\infty$ of G, *viz*, $G(\pm\infty, \tau) = 0$, insure proper convergence of the right member of (20) at infinity. From the continuity of the kernel it is clear that $\mathfrak{X}^{(+)}\ (T, \tau) = \mathfrak{X}^{(-)}\ (0, \tau) = 0$, all τ in $(0, T)$. Thus, the approach at this point uses this continuity property, along with the fact that the kernel K_0, corresponding to (19a), can be represented as the sum of exponentials, to find these as yet unknown functions $\mathfrak{X}^{(\pm)}\ (t, \tau)$. When this is achieved, Fourier inversion of both members of (20) with respect to t yields the Fourier transform (F.T.) of $L_{10}\ (\tau, t)_T$, from which $L_{10}\ (\tau, t)$ (and $L_{10}\ (t, \tau)$) follow in turn. Paralleling the author's[26] steps, after taking the Fourier transform of both sides of (20), we write after a little manipulation

$$L_{10}(\tau, t)_T = 2 \int_{-\infty i}^{\infty i} \frac{e^{pt}\, dp}{2\pi i W_0(p/2\pi i)} \cdot \Bigg[S_G(p, \tau) + \frac{G^{(+)}(\tau)e^{-pT}}{c^{(+)} + p} + \frac{G^{(-)}(\tau)}{c^{(-)} - p}$$
$$- \int_{T+}^{\infty} G(\tau_0, \tau)e^{-p\tau_0}\, d\tau_0 - \int_{-\infty}^{0-} G(\tau_0, \tau)e^{-p\tau_0}\, d\tau_0$$
$$+ X^{(+)}(p, \tau) + X^{(-)}(p, \tau) \Bigg], \quad 0- < t < T+, \quad (21)$$

$= 0$ elsewhere, (for t),

where by definition

$$S_G(p, \tau) = \int_{-\infty}^{\infty} e^{-p\tau_0} G(\tau_0, \tau)\, d\tau_0$$
$$= e^{-p\tau}[W_1(p/2\pi i) - W_0(p/2\pi i)]. \qquad (22)$$

Here $X^{(\pm)}\ (p, \tau)$ are the F.T.'s of $\mathfrak{X}^{(\pm)}\ (t, \tau)$,[27] with $p = i\omega = 2\pi i$, and by analytic continuation p is allowed at the appropriate stages to assume complex values.

Next, let us follow the direct procedure as indicated above,[23] to obtain the somewhat more useful form for explicit calculation [*cf.* (27), (28), below], using in place of (20) the "extended" form of G,[28] and carrying out the steps indicated by the author.[29] The result is

$$L_{10}(\tau, t)_T = 2 \int_{-\infty i}^{\infty i} \frac{e^{pt}\, dp}{2\pi i W_0(p/2\pi i)} \Bigg[\int_{0-}^{T+} G(\tau_0, \tau)e^{-p\tau_0}\, d\tau_0$$
$$+ \frac{G^{(+)}(\tau)e^{-pT}}{b_1^{(0)} + p} + \frac{G^{(-)}(\tau)}{b_1^{(0)} - p}$$
$$+ \mathfrak{X}^{(+)}(p, \tau) + \mathfrak{X}^{(-)}(p, \tau) \Bigg], \quad 0- < t < T+,$$
$$= 0, \qquad t < 0-, \qquad t > T+, \qquad (23)$$

[19] Middleton, *op. cit.*, Reference 2, Sect. A. 2.3.
[20] Rational spectra may be generated, for example, by passing iginally "white" noise through a linear, stable, invariant, lumped-nstant network. Thus, if

$$K_0(|\,t - \tau\,|) = \sum_{1}^{N_0} a_n^{(0)} \exp\,(-b_n^{(0)}\,|\,t - \tau\,|),$$

9a) is the associated spectral intensity density; resolution into rtial fractions determines the $a_n{}^{(0)}$ in terms of the $c_n{}^{(0)}$, and vice rsa.
[21] Middleton, *op. cit.*, Reference 2, Sect. A.2.1, p. 1082 *et seq.*
[22] *Ibid.*, footnote on p. 1083.
[23] *Ibid.*, Sect. A.2.3, pp. 1086–1091.
[24] *Ibid.*, Sect. A.2.4.
[25] *Ibid.*, (A.2–22)–(A.2–24).
[26] *Ibid.*, (A.2–25)–(A.2–31).
[27] *Ibid.*, (A.2–21, 27).
[28] *Ibid.*, (A.2–24).
[29] *Ibid.*, (A.2–25)–(A.2–43).

in which

$$\frac{\mathfrak{X}^{(\mp)}(p, \tau)}{\mathcal{W}_0(p/2\pi i)} = \sum_{n=2}^{N_0} \Gamma_n^{(\mp)}(\tau) \frac{(e^{-pT} \text{ or } 1)(b_1^{(0)} - b_n^{(0)})}{(b_n^{(0)} \pm p)(b_1^{(0)} \pm p)} \mathcal{W}_0(p/2\pi i)^{-1}. \quad (24)$$

Here we have set $c^{(+)} = b_1 = c^{(-)}$ for convenience.[30] The $2N_0 - 2$ undetermined functions of τ, $\Gamma_n^{(\mp)}(\tau)$, are found by introducing the solution (23) into the original integral equation (18) and treating the ensuing relation as an identity. When this is done, it is found that the $\Gamma_n^{(\mp)}(\tau)$ are linear functions of $G, \dot{G}, \ddot{G}, \cdots, G^{(2N_0-2M_0-1)}$, at $t = T$, 0, where differentiations (dots) are with respect to t.

Combining (24) and the above, we get

$$\begin{aligned} L_{10}(\tau, t)_T &= \int_{-\infty i}^{\infty i} \left[\frac{\mathcal{W}_1(p/2\pi i) - \mathcal{W}_0(p/2\pi i)}{\mathcal{W}_0(p/2\pi i)}\right] \frac{e^{p(t-\tau)}}{2\pi i}\, dp \\ &+ 2[K_1(T-\tau) - K_0(T-\tau)]e^{-b_1(t-T)} \big|_{t>T} \\ &+ 2[K_1(\tau) - K_0(\tau)]e^{b_1 t} \big|_{t<0} \\ &+ 2[K_1(t-\tau) - K_0(t-\tau)] \big|_{t>T, t<0} \\ &+ 2 \sum_{n=2}^{N_0} (b_1^{(0)} - b_n^{(0)}) \\ &\cdot \left\{\Gamma_n^{(+)}(\tau) \int_{-\infty i}^{\infty i} \frac{e^{-pT+pt}}{(b_n^{(0)} + p)(b_1^{(0)} + p)} \mathcal{W}_0(p/2\pi i)^{-1} \frac{dp}{2\pi i} \right. \\ &+ \left. \Gamma_n^{(-)}(\tau) \int_{-\infty i}^{\infty i} \frac{e^{pt}}{(b_n^{(0)} - p)(b_1^{(0)} - p)} \mathcal{W}_0(p/2\pi i)^{-1} \frac{dp}{2\pi i}\right\} \\ &\qquad 0- < t, \tau < T+, \quad (25) \end{aligned}$$

where $t-T$, etc. signifies that the preceding quantity is nonzero only for $t > T$, etc. Note that $L_{10}(\tau, t) = 0$ for all t outside $(0-, T+)$, but that $L_{10}(\tau, t)$ does not necessarily vanish for τ outside the square $(0- < t, \tau < T+)$. In fact, $L_{10}(\tau, t) \neq 0$, generally in the strip $(0- < t < T+)$, all τ, and, moreover, $L_{10}(\tau, t) \neq L_{10}(t, \tau)$, corresponding in the discrete, matrix forms (9c), to the fact that $\mathbf{K}_1\mathbf{K}_0^{-1} \neq \mathbf{K}_0^{-1}\mathbf{K}_1$ (unless $\mathbf{K}_0$ or $\mathbf{K}_1 = \mathbf{I}$).

An alternative form for (25) which is often more convenient for explicit computation may be obtained from (23). It is explicitly

$$\begin{aligned} L_{10}(\tau, t) &= 2 \int_{-\infty i}^{\infty i} \frac{e^{p(t-\tau)}}{\mathcal{W}_0(p/2\pi i)} \left\{K_1(t-\tau) - K_0(t-\tau) \right. \\ &+ \frac{e^{p(\tau-T)}}{b_1^{(0)} + p}[K_1(T-\tau) - K_0(T-\tau)] \\ &+ \frac{e^{p\tau}}{b_1^{(0)} - p}[K_1(\tau) - K_0(\tau)] \\ &+ \sum_{n=2}^{N_0} [b_1^{(0)} - b_n^{(0)}]\left(\Gamma_n^{(+)}(\tau) \frac{e^{p(\tau-T)}}{(b_n^{(0)} + p)(b_1^{(0)} + p)} \right. \\ &+ \left.\left. \Gamma_n^{(-)}(\tau) \frac{e^{p\tau}}{(b_n^{(0)} - p)(b_1^{(0)} - p)}\right)\right\} \frac{dp}{2\pi i} \quad (26) \end{aligned}$$

for $0- < t, \tau < T+$, with $L_{10}(\tau, t) = 0$, when t is outside $(0-, T+)$.

As specific examples, we see at once from the same source,[31] in the case of RC spectra, that

$$\begin{aligned} L_{10}(\tau, t)_{RC} &= \int_{-\infty i}^{\infty i} \left[\frac{\mathcal{W}_1(p/2\pi i) - \mathcal{W}_0(p/2\pi i)}{\mathcal{W}_0(p/2\pi i)}\right]_{RC} e^{p(t-\tau)} \frac{dp}{2\pi i} \\ &+ [\dot{K}_1(T-\tau) - \dot{K}_0(T-\tau) + \alpha_0 K_1(T-\tau) \\ &- \alpha_0 K_0(T-\tau)]\, \delta(t-T) + [-\dot{K}_1(\tau) + \dot{K}_0(\tau) \\ &+ \alpha_0 K_1(\tau) - \alpha_0 K_0(\tau)]\, \delta(t-0), \quad 0- < t < T+, \quad (27) \end{aligned}$$

where specifically

$$K_0(t) = \psi_0 e^{-\alpha_0|t|}; \qquad K_1(t) = \psi_1 e^{-\alpha_1|t|};$$

$$\mathcal{W}_0 = \frac{4\psi_0\alpha_0}{\alpha_0^2 - p^2}; \qquad \mathcal{W}_1 = \frac{4\psi_1\alpha_1}{\alpha_1^2 - p^2}. \quad (27a)$$

Note that

$$\dot{K}_{1,0}(0) = \tfrac{1}{2} \lim_{\epsilon \to 0} [\dot{K}_{1,0}(0+) + \dot{K}_{1,0}(0-)] = 0$$

here; observe, also that the integral in (27) is simply the spectral equivalent of the time-form

$$(2\psi_0\alpha_0)^{-1}\left(\alpha_0^2 - \frac{d^2}{dt^2}\right)[K_1(t-\tau) - K_0(t-\tau)]_{RC}.$$

Similarly, we find directly from Middleton,[32] that for LRC spectra

$$\begin{aligned} &L_{10}(\tau, t)_{LRC} \\ &= \int_{-\infty i}^{\infty i} \left[\frac{\mathcal{W}_1(p/2\pi i) - \mathcal{W}_0(p/2\pi i)}{\mathcal{W}_0(p/2\pi i)}\right]_{LRC} e^{p(t-\tau)} \frac{dp}{2\pi i} \\ &+ [H^{(3)} + 4(\omega_F^{(0)2} - \omega_0^{(0)2})\dot{H} + 2\omega_F^{(0)}\omega_0^{(0)2}H]_{T-\tau}\, \delta(t-T) \\ &- [\ddot{H} + 2\omega_F^{(0)}\dot{H} + \omega_0^{(0)2}H]_{T-\tau}\, \delta'(t-T) \\ &+ [H^{(3)} - 4(\omega_F^{(0)2} - \omega_0^{(0)2})\dot{H} + 2\omega_F^{(0)}\omega_0^{(0)2}H]_\tau\, \delta(t-0) \\ &+ [\ddot{H} - 2\omega_F^{(0)}\dot{H} + \omega_0^{(0)}H]_\tau\, \delta'(t-0), \\ &\qquad 0- < t < T+, \quad (28) \end{aligned}$$

where now $H(x) \equiv K_1(x) - K_0(x)$, and

$$K_a(t)_{LRC} = \psi_a \exp(-\omega_F^{(a)}|t|) \cdot \left(\cos \omega_1^{(a)} t + \frac{\omega_F^{(a)}}{\omega_1^{(a)}} \sin \omega_1^{(a)} |t|\right) \quad (28a)$$

$$\mathcal{W}_a(p/2\pi i)_{LRC} = W_0^{(a)}\omega_0^{(a)4}/(\omega_0^{(a)4} - 2(2\omega_F^{(a)2} - \omega_0^{(a)2})p^2 + p^4), \quad (28b)$$

with $W_0^{(a)} = 8\psi_a\omega_F^{(a)}/\omega_0^{(a)2}$, $\omega_1^2 = \omega_0^2 - \omega_F^2$ etc. In both cases, of course, $L_{10}(\tau, t) = 0$, t outside $(0-, T+)$. As in the RC case above, we readily find that the integral

[30] *Ibid.*, the comment following (A.2–38) on p. 1089.

[31] *Ibid.*, (A.2–48).

[32] *Ibid.*, (A.2–19), (A.2–54), (A.2–55).

(28) is again the spectral equivalent of the time-form

$$\frac{2}{{}^{(0)}\omega_0^{(0)^4}}\left[\omega_0^{(0)^4} - 2(2\omega_F^{(0)^2} - \omega_0^{(0)^2})\frac{d^2}{dt^2} + \frac{d^4}{dt^4}\right] \cdot [K_1(t-\tau) - K_0(t-\tau)]_{\mathrm{LRC}}.$$

We are now ready to compute $\int_{0-}^{T+} L_{10}(t, t)dt$ as required (1), (14), and (17) *et seq.* for the *n.* and *s.* condition of gularity. Setting $\tau = t$ in (25), we have directly

$$\int^{T+} L_{10}(t, t)\, dt = 2T \int_0^\infty \left[\frac{\mathcal{W}_1(f)}{\mathcal{W}_0(f)} - 1\right] df + \{\text{a finite number of finite terms}\}. \quad (29a)$$

he unspecified finite terms referred to in (29a) are the sult of integrating (with $\tau = t$) the sum in (25) and serving that typical integrals are of the form

$$\int_{0-}^{T+} K_{1,0}^{(A)}(T-t)\ \delta^{(B)}(t-T)\, dt,$$

$$\int_{0-}^{T+} K_{1,0}^{(A)}(t)\ \delta^{(B)}(t-0)\, dt,$$

here $A, B = 0, 1, \cdots, 2N_0 - 2M_0 - 1, 0 \le A + B \le$ $N_0 - 2M_0 - 1$, with various combinations of lower-rder derivatives of $K_{1,0}$ and the delta functions.[33] ntegration always yields a finite result, since $K_{1,0}^{(C)}(t)$ $\sum_n (d^C/dt^C)\ a_n e^{-b_n|t|}$ possesses all derivatives for $\ge 0+$, and, in particular, for $t \to 0+$ here (C = some teger or 0), while $K_{1,0}^{(C)}(T+)$ is also always bounded. Moreover, derivatives of $K_{1,0}$ never occur here to an order gher than $2N_0 - 2M_0 - 1$, and $K_0^{(2N_0-2M_0-1)}(0)$ is nsequently always finite [*cf.* the RC and LRC examples bove, for instance]. A similar remark can be made of ${}_1^{(2N_0-2M_0-1)}(0)$ if we observe that convergence of the tegral in (29a) implies that $\mathcal{W}_1/\mathcal{W}_0 = 1 + 0(p^{-2})$, and ence that $2N_1 - 2M_1 = 2N_0 - 2M_0$. In addition, the unspecified" terms in (29a) are always $0(p^{-1})$ in the tegrand *vis-à-vis* the leading or spectral-ratio terms, as amination of (21) shows, so that the latter dominate. hus, although $L_{10}(t, t)$ will usually contain delta functions[34] and their derivatives to an appropriate order at e two points $t = 0, T$, their contributions to the integral the left member of (29a) are finite, and consequently is the spectral term that establishes the convergence of is integral, and hence of the statistical test, by (1)–(3). precisely similar argument, starting instead with $= 0, b = 1$ in (18), yields

$$\int^{T+} L_{01}(t, t)\, dt = 2T \int_0^\infty \left[1 - \frac{\mathcal{W}_0(f)}{\mathcal{W}_1(f)}\right] df + \{\text{a finite number of finite terms}\}, \quad (29b)$$

hich may be combined with (29a) to give the composite rm (5a) of the *n.* and *s.* condition for these rational spectra. If, in addition, we require that these rational spectra[35] in the ratio $\mathcal{W}_1/\mathcal{W}_0$ (and $\mathcal{W}_0/\mathcal{W}_1$) are finite and nonzero[35] for all ($0 \le f < \infty$) we can then drop the integrals and express the *n.* and *s.* condition in its equivalent limit form (5b) above. Finally, we observe that the $\Gamma_n^{(\pm)}$ that appear in the unspecified finite terms referred to in (29a), *cf.* (25), are, of course, finite themselves, by the argument following (29a).

The proof of Theorem 2(b), for the singularity of the Bayes test in the case of rational spectra, is at once established by the same argument used in proving regularity above, where now the appropriate conditions of convergence are replaced by divergence. It remains only to show that (29a), (29b) in additive combination, *i.e.*, (6a) here, is the controlling expression *vis-à-vis* the finite number of now possibly divergent terms in the sums over n in (25) for both L_{10} and L_{01}. From (21)–(25) we see that the terms in $\mathfrak{X}^{(\pm)}(p)$, and hence the "unspecified" terms in (29), once more are $0(p^{-1})$ in the integrand *vis-à-vis* the leading or spectral-ratio terms in these expressions. Consequently, the singularities of these leading or spectral-ratio terms in combination are always of a different and higher order than those of the "unspecified" terms when f(or p) $\to \infty$, and are indeed the controlling terms, as required. For example, if $\mathcal{W}_1/\mathcal{W}_0 - 1$ is $0(f^2)$ as $f \to \infty$, then the "unspecified" terms are $0(f)$. Also, $1 - \mathcal{W}_0/\mathcal{W}_1$, is therefore $0(f^0)$, with its associated "unspecified" terms $0(f^{-1})$. When these two sets of terms are added, the divergent spectral ratio terms are $0(f^2)$ *vis-à-vis* $0(f)$ of the "unspecified" terms. This completes the demonstration of Theorems 2(a) and 2(b).

V. Nonrational Spectra; Special Cases

There remains the question of nonrational (nonband-limited) spectra in the stationary cases. Theorems 1(a) and 1(b) apply here, of course, but do not explicitly indicate the dependence on the spectral distribution. In the situation where nonrational spectra can be regarded as limiting forms of rational spectra, *cf.* (19) and for which the "end-effect" terms [*i.e.*, those involving $\Gamma_n^{(\pm)}$, *cf.* (24), (25)] also yield a finite contribution to the integral as in (29), the relation (5a) applies, by the argument of Section IV above for regularity in the rational cases. Similarly, (6a) applies for singularity, and Theorem 3, Section II, is the result.

Although we are not able at present to establish that (5a) and (6a) are the necessary and sufficient conditions for all nonrational spectra, we conjecture that this is indeed true, for on physical grounds alone we expect this to be the case, since nonrational spectra, for example, with "holes" on a finite frequency interval will lead to singular tests, *cf.* the remarks in 1) of Section VII below; while for

[33] *Ibid.*, the comments on p. 1091.
[34] An exception is the singular case where either N_1 or N_0 is "white".

[35] We remark again that the situations of multiple poles and zeros may be included here by appropriate limits on the coefficients $a_n^{(1,0)}$, $b_n^{(1,0)}$, etc., so that our results apply for all (finite intensity) rational spectra; *cf.* Middleton, *op. cit.*, Reference 2, (A. 2–58) and comments.

regularity the integral must at least be bounded. The main difficulty in proving the conjecture lies in showing that the terms in $\mathfrak{X}^{(*)}(p)$ here,[36] *cf.* (21)–(25), are indeed suitably finite on the one hand, and have lower-order singularities than the spectral term on the other hand, analogous to the behavior in the rational cases considered above (Section IV).

VI. Bandlimited Spectra: Proof of Theorem 4

As a final example of nonrational spectra, let us consider bandlimited processes[37] on $(0 < f < B)$. Here we outline an alternative proof to Slepian's earlier result,[4] *viz.*, that a sufficient condition that the Bayes test of two stationary gauss processes, N_1 vs N_0, be singular is that one (or both) of the processes be bandlimited. This is easily shown from the result of Theorem 1(b), *viz.*, (4) and the demonstration in Section IV, for from the author's relation[38] we may write the various covariance functions K_1, K_0, $K_{10} \equiv K_1 - K_0$ of these bandlimited processes as

$$K_{1,0,10}(\tau) = \sum_{-\infty}^{\infty} K_{1,0,10}(k/2B) \frac{\sin 2\pi B(\tau - k/2B)}{2\pi B(\tau - k/2B)} \tag{30}$$

and insert this into the basic integral equations for L_{10}, L_{01}, *cf.* (2). We observe directly that the solutions of (2) are then of the form $a_k\delta(t - u)$, all k, so that the integral $\int_{0-}^{T+} L_{ab}(t, t)dt \to \infty$, and consequently, that bandlimited processes lead to singular Bayes tests. Bandlimiting is clearly a *sufficient* condition, but not a necessary one, for singularity can also occur when the noise processes possess rational spectra, *cf.* Theorem 2(b).

VII. Concluding Remarks

Some of the principal implications of the preceding analysis may now be briefly summarized:

1) We observe first that the mathematically significant models of physical situations (involving the optimum tests above) are nonsingular. The singular model, if it is chosen or constructed, is *not* an adequate, or even acceptable representation, from an applied point of view, since it leads to physically unrealizable outcomes. As an example, let us suppose that one of the noise processes in our treatment above has a finite spectral gap on some frequency interval where the other process does not. Then it is clear that we would expect a perfect test of H_1 vs H_0, as confirmed by condition (6a), simply by using a filter located in the gap and observing its zero or nonzero output, as N_1 or N_0 occurs. Similar examples can be constructed: for example, if $\mathcal{W}_1/\mathcal{W}_0$ does not approach unity for $f \to \infty$, or if $\mathcal{W}_1\mathcal{W}_0 = 0$ at some $f = f_0$, etc., *cf.* (6b).

2) It is *not* necessarily enough, however, that a test be regular (*i.e.*, nonsingular) for it to be an adequate representation of a physical situation. This is indicated by the following example: Consider the FSK situation, where sinusoidal signals, transmitted through a scattering medium, are received as narrow-band normal processes, *e.g.*, fast, normal fading, N_0, N_1 above. Now suppose that each FSK and corresponding scattered signal, is of equal intensity, respectively, and that their spectra are identical except for location, or in any case are such that (5) holds. Detection of N_1 vs N_0 is accordingly nonsingular. But, let us change the level of one of the transmitted sinusoids; then $\lim_{f\to\infty} \mathcal{W}_1/\mathcal{W}_0 \neq 1$, and we have a singular situation, which according to our model can occur for the slightest difference in level here. Physically, of course, we know that perfect discrimination does not occur under these circumstances; we are at least ultimately limited by the inherent noisiness of the observation process itself, here generated by the receiver. An acceptable model adds a background noise $N(t)$; thus, $N_1 = S_2 + N$; $N_0 = S_1 + N$, where S_1, S_2 are the received FSK noise signals. In this condition, an equivalent statement of (5a) is for rational spectra and for certain limiting classes of irrational spectra

$$\left| 2T \int_0^{\infty} \left[\frac{\mathcal{W}_{S_2}(f) - \mathcal{W}_{S_1}(f)}{\mathcal{W}_N(f)} \right] df \right| < \infty. \tag{31}$$

Similarly, in the "on-off" problem analyzed earlier by the author,[1] where $N_1 = S + N_0$; $N_0 = N_0$, the n. s. condition (5a) may alternatively be expressed as

$$2T \int_0^{\infty} \frac{\mathcal{W}_S(f)}{\mathcal{W}_N(f)} df < \infty. \tag{32}$$

Both (31) and (32) require a suitably rapid fall-off of $\mathcal{W}_S(f)$ *vis-à-vis* $\mathcal{W}_N(f)$ as $f \to \infty$, *e.g.*, $\mathcal{W}_S/\mathcal{W}_N = 0(f^{-1-\epsilon})$ which in the rational cases is $0(f^2)$ at least (and that $\mathcal{W}_S/\mathcal{W}_N$ is bounded and nonzero for $0 \leq f < \infty$).

3) The above suggests that a sufficient condition for "stable" regularity, and hence an acceptable model, is the addition of a suitable background noise, in physical situations a thermal process of some kind, generated either internally or externally to the receiver, whose spectral extent and behavior as $f \to \infty$ always insure the conditions (31), or (32), or (5).

We remark that other possibilities insuring regularity may also exist. For example, if the uncertainty with which spectral shape is measurable is significant compared to the perturbing effects of an additive background noise, this uncertainty, represented analytically by one or more random parameters, *e.g.*, spectral width, fall-off at high frequencies, or other "shape-factors", may be enough to establish convergence of the test, with, of course, appropriate averages over these random parameters in the optimum structure (7), *cf.* the analogous situation with deterministic signals.[39] In any case, the appropriate approach is the one which incorporates the dominant uncertainty in the physical model.

[36] The sum over n in (25) is replaced by two unknown functions of t, τ, while the other terms remain unchanged.

[37] These include processes with one or more bandlimited component processes, as well.

[38] Middleton, *op. cit.*, Reference 2, (4.37).

[39] *Ibid.*, (19.20), (20.1).

4) In the "stable" regular cases, then, the exact spectral stributions for large frequencies are not critical; dection is essentially a distinction between *energy* states, d it is not the detailed spectral shape that is controlling, ovided, of course, that conditions like (5), (31), (32) are tisfied. This is what allows us to apply with success in actice our analytical models admittedly imprecise in tail.[40] Of course, an optimum system attempts to match" itself suitably[41] to the incoming waves: broad ectra require broad filters, etc., a fact taken into account radiometry theory and practice, for example, where the esign trend is toward increasingly broad filters, using avelling wave amplifiers, etc., to match the radio sources nder study, themselves spectrally very broad. For given rror probabilities in detection, the result is a correspondng reduction in effective data acquisition time: the roader the spectrum the shorter the period one needs to nake an observation at a given error. In this sense, we night say that here design based on regularity, in taking nto account spectral behavior at high frequencies, pproaches the limiting form of a singular test, which, of ourse, it can never reach for the reasons cited above.

5) In the nonsingular cases, the passage from n sampled values of the process on $(0, T)$ to the continuous limit as $n \to \infty$ is valid for all adequate models. All pertinent nformation concerning the process is efficiently employed, nd none is thrown away (up to the point of an actual lecision).

6) We conjecture that the conditions (5), (6), are also oth necessary and sufficient for regularity or singularity, respectively, for all nonrational (nonbandlimited) spectra, although we have been able to establish this for special cases of nonrational spectra only, *cf.* comments in Section V and VI.

[40] *Ibid.*, the comments on pp. 825, 826.
[41] *Ibid.*, Sect. 20.2–4.

7) Since bandlimiting is a sufficient condition for singularity in the stationary cases (Theorem 4), we might think it possible always to insure perfect detection with arbitrarily small samples simply by introducing an ideal band-pass filter ($0 < f < B$) at the front end of the Bayes receiver. Physically, of course, such filters in lumped-constant or distributed form cannot be made to be noise-free, so that a nonspectrally limited residual additive noise always accompanies the bandlimited input to the Bayes receiver, and the situation decribed in (3) is once more in force. Alternatively, if the ideal band-pass filter is represented by a suitable computer, it may be possible to eliminate the effects of inherent noise, but only at the expense of an infinitely long processing time, with the practical effect that singularity again cannot be achieved by bandlimiting in physical situations.

8) By similar arguments, we expect the same general conclusions to apply for non-Gaussian processes, although the conditions for regularity and singularity are predicted to be much more involved, since the statistical description of such processes is likewise much more complicated than for the Gaussian process.

Acknowledgment

The author is indebted to Dr. T. S. Pitcher, Lincoln Laboratory, and Dr. P. Bello, Applied Research Laboratory, Sylvania Electronic Systems, for valuable comments and criticisms.

Correspondence

A Lower Bound for Error-Detecting and Error-Correcting Codes*

The purpose of this note is to establish a new lower bound implicit in Theorems 1, 2, and 3 below for error-detecting and error-correcting codes.

Notations, Definitions, and Previous Theorems

Let G_n denote the set of all binary representations of n digits. Under digit-wise modulo 2 addition, G_n is a group of 2^n elements. Furthermore, by defining the distance $D(x, y)$ between elements x and y in G_n to be the number of digits where x and y are different, G_n is a metric space.

Using this metric, Hamming showed in 1950[1] that a subset G of G_n forms a set of k error-detecting codes if G is such that x, y in G implies that $D(x, y) \geq 2k$. If $D(x, y) \geq 2k + 1$, for x, y in G, then G is a set of k error-correcting codes. Furthermore, G is a group code if G has the additional property of being a subgroup of G_n.

The number of elements in G is bounded above by the following limits established also by Hamming. Let $B(n, 2k + 1)$ be the maximum number of k error-correcting codes in G_n, and $B(n, 2k)$ the maximum number of k error-detecting codes in G_n; then the following upper bounds hold:

$$B(n, 1) = 2^n$$

$$B(n, 2) = 2^{n-1}$$

$$B(n, 3) = 2^m \leq 2^n/(n + 1)$$

$$B(n, 4) = 2^m \leq 2^{n-1}/n$$

$$B(n - 1, 2k - 1) = B(n, 2k)$$

$$B(n, 2k - 1) = 2^m \leq 2^n/[1 + C(n, 1) + \cdots C(n, k - 1)],$$

where m is an integer and denotes the number of digits that may be used for information bits, and where $C(n, p)$ denotes the binomial coefficient $\binom{n}{p}$; that is,

$$C(n, p) = n!/[(n - p)!p!].$$

In 1959, Shapiro and Slotnick[2] presented two results on lower bounds. They are

$$B(n, k) \geq 2^n/[1 + C(n, 1) + \cdots + C(n, k - 1)], \quad (1)$$

and, for an infinite sequence of n,

$$B(n, k) \geq 2^n/[1 + C(n, 1) + \cdots + C(n, k - 2)]. \quad (2)$$

A New Lower Bound

With this as background, I should like to establish a new lower bound for error-detecting and error-correcting codes.

Lemma 1

For fixed integers α and k, there exists N such that whenever $n \geq N$,

$$\sum_{i=0}^{k-1} C(n - \alpha, i) \geq 2 \sum_{i=0}^{k-2} C(n, i).$$

Proof

Express

$$\sum_{i=0}^{k-1} C(n - \alpha, i) - 2 \sum_{i=0}^{k-2} C(n, i)$$

as a polynomial:

$$a_{k-1}n^{k-1} + a_{k-2}n^{k-2} + \cdots + a_1 n + a_0.$$

It is clear that $a_{k-1} > 0$.

Lemma 1 now follows as a consequence of the fact that if a polynomial in n has positive first coefficient, it is greater than zero for n large. Q.E.D.

Lemma 2

For group codes,[3] if

$$B(n, k) < \frac{2^n}{1 + C(n, 1) + \cdots C(n, k - 2)},$$

then

$$B(n + 1, k) = 2B(n, k).$$

Proof

See Shapiro and Slotnick.[2]

Theorem 1

For group codes,

$$B(n, k) \geq \frac{2^n}{1 + C(n - 1, 1) + \cdots C(n - 1, k - 1)}$$

for n sufficiently large.

Proof

By Lemma 1, we can choose for fixed k, an N such that whenever $n \geq N$,

$$\sum_{i=0}^{k-1} C(n - 1, i) \geq 2 \sum_{i=0}^{k-2} C(n - 1, i).$$

* Received by the PGIT, June 13, 1960; revised manuscript received, August 4, 1960.

[1] R. W. Hamming, "Error-detecting and error-correcting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160; April, 1950.

[2] H. S. Shapiro, D. L. Slotnick, "On the mathematical theory of error-correcting codes," *IBM J. Res. Dev.*, vol. 3, pp. 25–34; January, 1959.

[3] For a discussion of the properties of group codes, refer to D. Slepian, "A class of binary signaling alphabets," *Bell Sys. Tech. J.*, vol. 35, pp. 203–233; January, 1956.

Suppose that

$$B(n-1, k) < \frac{2^{n-1}}{1 + C(n-1, 1) + \cdots C(n-1, k-2)};$$

then, by Lemma 2,

$$B(n, k) = 2B(n-1, k).$$

Using Shapiro and Slotnick's lower bound, we have (1) above,

$$B(n-1, k) \geq \frac{2^{n-1}}{1 + C(n-1, 1) + \cdots C(n-1, k-1)};$$

therefore,

$$B(n, k) = 2B(n-1, k) \geq \frac{2^n}{1 + C(n-1, 1) + \cdots C(n-1, k-1)}.$$

Suppose that

$$B(n-1, k) \geq \frac{2^{n-1}}{1 + C(n-1, 1) + \cdots C(n-1, k-2)},$$

then

$$B(n, k) \geq B(n-1, k) \geq \frac{2^{n-1}}{1 + C(n-1, 1) + \cdots C(n-1, k-2)}$$

$$= \frac{2^n}{2 \sum_{i=0}^{k-2} C(n-1, i)} \geq \frac{2^n}{\sum_{i=0}^{k-1} C(n-1, i)}$$

whenever $n \geq N$. Q.E.D.

Theorem 2

For fixed integers α and k, there exists N such that whenever $n \geq N$

$$B(n, k) \geq \frac{2^n}{1 + C(n-\alpha, 1) + \cdots C(n-\alpha, k-1)}$$

for group codes.

Proof

By induction.

Let k be a fixed integer.

Let $\alpha = 1$. Theorem 1 shows that the above statement holds.

Assume that the theorem holds for $\alpha = \bar{\alpha}$. Then it suffices to show that the theorem holds for $\alpha = \bar{\alpha} + 1$.

Let $N_{\bar{\alpha}}$ be the number such that whenever $n \geq N_{\bar{\alpha}}$,

$$B(n, k) \geq \frac{2^n}{1 + C(n-\bar{\alpha}, 1) + \cdots C(n-\bar{\alpha}, k-1)}.$$

Let $N_{\bar{\alpha}+1} \geq N_{\bar{\alpha}} + 1$, and

$$\sum_{i=0}^{k-1} C(n-\bar{\alpha}-1, i) \geq 2 \sum_{i=0}^{k-2} C(n-1, i)$$

when $n \geq N_{\bar{\alpha}+1}$. (We can choose such a $N_{\bar{\alpha}+1}$ due to Lemma 1.) Let $n \geq N_{\bar{\alpha}+1}$.

Case 1

Suppose

$$B(n-1, k) < \frac{2^{n-1}}{1 + C(n-1, 1) + \cdots C(n-1, k-2)};$$

then

$$B(n, k) = 2B(n-1, k) \geq \frac{2^n}{1 + C(n - \bar{\alpha} - 1, 1) + \cdots C(n - \bar{\alpha} - 1, k-1)}$$

$$= \frac{2^n}{1 + C(n - (\bar{\alpha} + 1), 1) + \cdots C(n - (\bar{\alpha} + 1), k-1)}.$$

Case 2

Suppose

$$B(n-1, k) \geq \frac{2^{n-1}}{1 + C(n-1, 1) + \cdots C(n-1, k-2)};$$

then

$$B(n, k) \geq B(n-1, k) \geq \frac{2^n}{2 \sum_{i=0}^{k-2} C(n-1, i)}$$

$$\geq \frac{2^n}{1 + C(n - (\bar{\alpha} + 1), 1) + \cdots C(n - (\bar{\alpha} + 1), k-1)}.$$

Q.E.D.

Lemma 3

If

$$\sum_{i=0}^{k-1} C(n - \alpha, i) \geq 2 \sum_{i=0}^{k-2} C(n, i),$$

then

$$\sum_{i=0}^{k-1} C(n + 1 - \alpha, i) \geq 2 \sum_{i=0}^{k-2} C(n + 1, i).$$

Proof

$$\sum_{i=0}^{k-1} C(n + 1 - \alpha, i) = \sum_{i=0}^{k-1} C(n - \alpha, i) + \sum_{i=0}^{k-2} C(n - \alpha, i);$$

$$2 \sum_{i=0}^{k-2} C(n + 1, i) = 2 \sum_{i=0}^{k-2} C(n, i) + 2 \sum_{i=0}^{k-3} C(n, i).$$

Hence, it suffices to show that

$$\sum_{i=0}^{k-2} C(n - \alpha, i) \geq 2 \sum_{i=0}^{k-3} C(n, i).$$

That is, we want to show that

$$\frac{\sum_{i=0}^{k-2} C(n - \alpha, i)}{2 \sum_{i=0}^{k-3} C(n, i)} \geq 1.$$

w,

$$\frac{\sum_{i=0}^{k-2} C(n-\alpha, i)}{2\sum_{i=0}^{k-3} C(n, i)} = \frac{\sum_{i=0}^{k-1} C(n-\alpha, i) - C(n-\alpha, k-1)}{2\sum_{i=0}^{k-2} C(n, i) - 2C(n, k-2)}.$$

t

$$p = \frac{C(n-\alpha, k-1)}{\sum_{i=0}^{k-1} C(n-\alpha, i)}; \; q = \frac{2C(n, k-2)}{2\sum_{i=0}^{k-2} C(n, i)};$$

en

$$\frac{\sum_{i=0}^{k-2} C(n-\alpha, i)}{2\sum_{i=0}^{k-3} C(n, i)} = \frac{(1-p)}{(1-q)} \cdot \frac{\sum_{i=0}^{k-1} C(n-\alpha, i)}{2\sum_{i=0}^{k-2} C(n, i)}.$$

et us compare p with q:

$$p < q, \quad \text{or} \quad \frac{C(n-\alpha, k-1)}{\sum_{i=0}^{k-1} C(n-\alpha, i)} < \frac{C(n, k-2)}{\sum_{i=0}^{k-2} C(n, i)},$$

rovided

$$\left[\frac{(n-\alpha)(n-1-\alpha)\cdots(n-k+2-\alpha)}{(k-1)!}\right] \cdot \left[1 + n + \frac{n(n-1)}{2!} + \cdots \frac{n(n-1)\cdots(n-k+3)}{(k-2)!}\right]$$

$$< \left[\frac{n(n-1)\cdots(n-k+3)}{(k-2)!}\right] \cdot \left[1 + (n-\alpha) + \frac{(n-\alpha)(n-1-\alpha)}{2!} + \cdots \frac{(n-\alpha)\cdots(n-k+2-\alpha)}{(k-1)!}\right].$$

In order to prove that the above in-
quality holds, let us express it as follows:

$$\left[\frac{(n-\alpha)(n-1-\alpha)\cdots(n-k+2-\alpha)}{(k-1)!}\right][A_1 + A_2 + \cdots + A_{k-2}]$$

$$< \left[\frac{n(n-1)\cdots(n-k+3)}{(k-2)!}\right] \cdot [B_1 + B_2 + \cdots + B_{k-2}],$$

here

$$A_1 = 1 \qquad B_1 = 1 + (n-\alpha)$$

$$A_2 = n \qquad B_2 = \frac{(n-\alpha)(n-1-\alpha)}{2!}$$

$$A_3 = \frac{n(n-1)}{2!} \qquad B_3 = \frac{(n-\alpha)(n-1-\alpha)(n-2-\alpha)}{3!}$$

$$\vdots$$

$$A_{k-2} = \frac{n(n-1)\cdots(n-k+3)}{(k-2)!} \quad B_{k-2} = \frac{(n-\alpha)\cdots(n-k+2-\alpha)}{(k-1)!}.$$

term-by-term comparison shows that
r each i, $i = 1, \cdots k-2$,

$$\frac{(n-\alpha)(n-1-\alpha)\cdots(n-k+2-\alpha)}{(k-1)!} \cdot A_i < \frac{n(n-1)\cdots(n-k+3)}{(k-2)!} \cdot B_i.$$

Thus, the inequality $p < q$ does in fact hold true.

Recall that by definition,

$$p < 1, \qquad q < 1.$$

Hence $p < q$ implies that

$$\frac{(1-p)}{(1-q)} > 1.$$

Therefore,

$$\frac{\sum_{i=0}^{k-1} C(n-\alpha, i)}{2 \sum_{i=0}^{k-2} C(n, i)} \geq 1$$

implies that

$$\frac{(1-p)}{(1-q)} \cdot \frac{\sum_{i=0}^{k-1} C(n-\alpha, i)}{2 \sum_{i=0}^{k-2} C(n, i)} \geq 1.$$

This gives us the desired result. Q.E.D.

Theorem 3

Given n, k, let α be largest integer such that

$$\sum_{i=0}^{k-1} C(n-\alpha, i) \geq 2 \sum_{i=0}^{k-2} C(n-1, i).$$

Then for group codes,

$$B(n, k) \geq \frac{2^n}{\sum_{i=0}^{k-1} C(n-\alpha, i)}.$$

Proof

This follows as an immediate consequence of Lemma 3, the previous theorems, and results of the proofs of the previous theorems.

Summary

Table I contains a comparison of the new lower bound with Shapiro and Slotnick's lower bound (1), and with Hamming's upper bound. The reader is reminded that the maximum number of k-error correcting or detecting codes is an integral power of 2 [*i.e.* $B(n, 2k) = 2^m$]. As is evident from the Table and from the statement of the theorems, the new results are considerably stronger than Shapiro and Slotnick's lower bound (1).

A comparison with Shapiro and Slotnick's lower bound (2) was not included in the Table since in this case the lower bound holds only for some unspecified infinite sequence. However, for any given n for which the lower bound (2) does hold, the result is a lower bound that is greater than results of the new theorems.

TABLE I

Magnitude of n, k	Shapiro and Slotnick's Lower Bound (1)	New Lower Bound	Hamming's Upper Bound
$n = 10$, $k = 4$	2^3	2^4	2^5
$n = 20$, $k = 5$	2^7	2^8	2^{12}
$n = 20$, $k = 4$	2^9	2^{12}	2^{14}
$n = 40$, $k = 5$	2^{24}	2^{26}	2^{30}
$n = 40$, $k = 4$	2^{27}	2^{30}	2^{33}
$n = 100$, $k = 5$	2^{79}	2^{82}	2^{87}
$n = 100$, $k = 4$	2^{83}	2^{87}	2^{92}
$n = 200$, $k = 7$	2^{164}	2^{168}	2^{179}

$k = 4 \Longrightarrow$ double error-detecting
$k = 5 \Longrightarrow$ double error-correcting

(Mrs.) Peggy Tang Strait
Research Associate
G. C. Dewey Corp.
New York, N. Y.

A Simple Proof of an Inequality of McMillan*

Let l_i, $i = 1, \cdots, b$, be the length of the ith word of a list of b words, each word being a string of letters from an alphabet of a letters. Assume that distinct strings of words from the list, when written out without additional space marks to separate the words, determine distinct strings of letters, so that the total number of strings of words of letter-length k is $\leqq a^k$. Let $l = \max_i l_i$. Then for all integers $n > 0$,

$$\left(\sum_{i=1}^{b} \frac{1}{a^{l_i}}\right)^n = \sum_{i_1, i_2, \cdots, i_n = 1}^{b} \frac{1}{a^{l_{i_1}} a^{l_{i_2}} \cdots a^{l_{i_n}}} = \sum_{j=1}^{b^n} \frac{1}{a^{L_j}},$$

where j runs over the b^n strings of n words and L_j is the number of letters in the jth string. Since $\max_j L_j = nl$ and $\min_j L_j \geqq n$, we have, grouping j's with $L_j = k$,

$$\left(\sum_{i=1}^{b} \frac{1}{a^{l_i}}\right)^n = \sum_{k=n}^{nl} \frac{\text{number of strings of } n \text{ words having } k \text{ letters}}{a^k}$$

$$\leqq \sum_{k=n}^{nl} \frac{a^k}{a^k} \leqq nl.$$

Since $x > 1$ implies $x^n > nl$ for sufficiently large n, it follows that $\sum_{i=1}^{b} a^{-l_i} \leqq 1$, which is McMillan's[1] result.

Jack Karush
Dept. of Statistics
University of California
Berkeley, Calif.

* Received by the PGIT, October 17, 1960. This note was prepared with the partial support of the Office of Ordnance Research, U. S. Army, under Contract DA-04-200-ORD-171, Task Order 3.

[1] B. McMillan, "Two inequalities implied by unique decipherability," IRE Trans. on Information Theory, vol. IT-2, pp. 115–116; December, 1956.

Note on an Integral Equation Occurring in the Prediction, Detection, and Analysis of Multiple Time Series*

The Wiener-Hopf integral equation with a finite upper limit

$$\int_0^T W(\tau) R(t-\tau)\, d\tau = f(t) \qquad (0 \leq t \leq T), \qquad (1)$$

where $R(\tau)$ is the correlation function of a stationary process, is frequently encountered in the theories of prediction and detection and in the analysis of stationary time series. About four years ago, the authors encountered a matrix form of this equation in the course of attempting to determine the amount of information contained in a finite segment of a stationary Gaussian process about a finite segment of another stationary Gaussian process, with the two processes stationarily correlated with one another. This problem was subsequently solved by Gel'fand and Yaglom.[1]

In working on this problem we have obtained as byproducts a vector form of the Karhunen-Loeve expansion and an extension of Preston's generalized probability-density functional[2] to vector Gaussian processes. Some of these results were applied by one of the authors to the solution of optimal reception problems involving the processing of $n(n \geq 1)$ signals.[3]

The purpose of the present note is to give a brief account of the foregoing results and, more particularly, to indicate a general method of solving matrix equations of the form (1) for the case where the elements of the spectral density matrix $\mathbf{G}(\omega)$ are rational functions of ω. Such equations seem to play a basic role in the prediction, detection, and analysis of multiple time series.

Consider the matrix integral equation

$$\int_0^T \mathbf{R}(t-\tau) \mathbf{W}(\tau)\, d\tau = \mathbf{f}(t) \qquad (0 \leq t \leq T), \qquad (2)$$

* Received by the PGIT, July 1, 1960; revised manuscript received, September 29, 1960. This work was supported in part by the National Science Foundation.

[1] I. M. Gel'fand and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function," *Uspekhi Mat. Nauk*, vol. 12, no. 1, pp. 3–52; January, 1957. (In Russian.)

[2] G. W. Preston, "The equivalence of optimum transducers and sufficient and most efficient statistics," *J. Appl. Phys.*, vol. 24, pp. 841–844; July, 1953.

[3] J. B. Thomas and E. Wong, "On the statistical theory of optimum demodulation," IRE Trans. on Information Theory, vol. IT-6, pp. 420–425; September, 1960.

[wh]ere $\mathbf{R}(\tau)$ is an $n \times n$ correlation matrix, [an]d $\mathbf{W}$ and $\mathbf{f}$ are n-vectors. Let $\mathbf{R}^{-1}(t, \tau)$ [de]note a matrix kernel inverse to $\mathbf{R}(t - \tau)$ [in] the sense that

$$\int^T \mathbf{R}(t - \tau)\mathbf{R}^{-1}(\tau, \xi)\, d\tau = \mathbf{I}\, \delta(t - \xi), \quad (0 \leq t, \xi \leq T), \quad (3)$$

[wh]ere $\mathbf{I}$ is the identity matrix. Then the [so]lution of (2) can be written by super[po]sition as

$$[\mathbf{W}](\tau) = \int_{-\infty}^{\infty} 1(\xi)1(T - \xi) \cdot \mathbf{R}^{-1}(\tau, \xi)\mathbf{f}(\xi)\, d\xi, \quad (4)$$

[w]here the unit-step functions serve to [li]mit the range of integration to the interval [0], T]. This mode of limiting the range of [in]tegration serves the purpose of resolving [th]e ambiguity arising when $R^{-1}(\tau, \xi)$ con[ta]ins delta functions at the points $\xi = 0$ [a]nd $\xi = T$.

By virtue of (4), the problem of solving [(]2) reduces to that of solving the matrix [in]tegral equation (3). Before describing a [m]ethod of solving the latter, we shall [b]riefly indicate the manner in which this [e]quation arises in the expansion of vector [p]rocesses and in the generalized probability-[d]ensity functional for Gaussian processes.

For simplicity, we shall restrict the dis[c]ussion to the case of two stationary and [s]tationarily-correlated second-order proc[e]sses, $\{x(t)\}$ and $\{y(t)\}$, having zero means. [F]or such processes, it can readily be shown [t]hat $x(t)$ and $y(t)$ admit of representation [(]with convergence in quadratic mean) as

$$x(t) = \sum_{\mu=1}^{\infty} \alpha_\mu \varphi_\mu(t), \quad (5)$$

$$y(t) = \sum_{\mu=1}^{\infty} \alpha_\mu \theta_\mu(t), \quad (6)$$

[w]here the φ_μ and θ_μ are orthogonal in the [s]ense that

$$\int_0^T [\varphi_\mu(t)\varphi_\nu(t) + \theta_\mu(t)\theta_\nu(t)]\, dt = \delta_{\mu\nu} \quad (\mu, \nu = 1, 2, 3, \cdots), \quad (7)$$

[a]nd the α_μ are random variables given by

$$[\alpha]_\mu = \int_0^T [x(t)\varphi_\mu(t) + y(t)\theta_\mu(t)]\, dt. \quad (8)$$

Furthermore, the φ_μ and θ_μ are eigen[f]unctions of the system of integral equations

$$[\varphi]_\mu(t) = \lambda_\mu \int_0^T [\varphi_\mu(\xi)R_{xx}(t - \xi) + \theta_\mu(\xi)R_{xy}(t - \xi)]\, d\xi,$$

$$[\theta]_\nu(t) = \lambda_\nu \int_0^T [\varphi_\nu(\xi)R_{yx}(t - \xi) + \theta_\nu(\xi)R_{yy}(t - \xi)]\, d\xi, \quad (9)$$

[w]ith $E\{\alpha_\mu \alpha_\nu\} = \lambda_\mu^{-1}\delta_{\mu\nu}$. The functions [$R$]$_{xx}$, R_{xy}, R_{yx} and R_{yy} in (9) are the elements of the correlation matrix of $\{x(t)\}$ and $\{y(t)\}$. Thus, the expansion of $\{x(t)\}$ and $\{y(t)\}$ in the form of (5) and (6) reduces to solving a homogeneous form of the matrix integral equation (2).

Next, assume that $\{x(t)\}$ and $\{y(t)\}$ are jointly Gaussian processes. Consider a strip B_x of width $2h$ centering on a curve $u(t)$, $0 \leq t \leq T$, in the (x, t) plane, and let B_y be a strip of width $2h'$ centering on a curve $v(t)$, $0 \leq t \leq T$, in the (y, t) plane. Similarly, let B_x^0 and B_y^0 be strips of widths $2h$ and $2h'$ centering on the t axis in the (x, t) and (y, t) planes, respectively. Then the generalized joint-probability-density functional for the processes $\{x(t)\}$ and $\{y(t)\}$ over the interval $[0, T]$ is given by[4]

$$P_T(u, v) = \lim_{h, h' \to 0} \frac{\Pr\{x(t) \in B_x \text{ and } y(t) \in B_y \text{ for } 0 \leq t \leq T\}}{\Pr\{x(t) \in B_x^0 \text{ and } y(t) \in B_y^0 \text{ for } 0 \leq t \leq T\}} \quad (10)$$

$$= \exp\left\{-\frac{1}{2}\int_0^T \int_0^T \mathbf{z}'(\tau)\mathbf{R}^{-1}(\tau, \xi)\mathbf{z}(\xi)\, d\tau\, d\xi\right\}, \quad (11)$$

where $\mathbf{z}(\tau)$ is a vector with components $u(\tau)$ and $v(\tau)$, and $\mathbf{z}'(\xi)$ is the transpose of $\mathbf{z}(\xi)$. Joint density functionals of this form can be employed effectively in problems involving Gaussian signals over finite intervals of time. Note that to find $P_T(u, v)$ one has to solve the integral equation (3) for $\mathbf{R}^{-1}(\tau, \xi)$.

We now proceed to outline a method of solution of this equation for the case where the spectral-density matrix $\mathbf{G}(\omega)$ is rational in ω. Without loss is generality, $\mathbf{G}(\omega)$ can be written as

$$\mathbf{G}(\omega) = \frac{1}{Q(j\omega)}\begin{bmatrix} N_{xx}(j\omega) & N_{xy}(j\omega) \\ N_{yx}(j\omega) & N_{yy}(j\omega) \end{bmatrix}, \quad (12)$$

where $Q(j\omega)$, $N_{xx}(j\omega)$, $\cdots$, $N_{yy}(j\omega)$ are polynomials in $j\omega$. Let the elements of $\mathbf{R}^{-1}(t, \tau)$ be denoted by $r_{xx}(t, \tau)$, $r_{xy}(t, \tau)$, $r_{yx}(t, \tau)$, $r_{yy}(t, \tau)$. Then, upon operating on both sides of (3) with the differential operator $Q(p)$, $p = d/dt$, $0 \leq t, \tau \leq T$, we note that a typical term such as

$$\int_0^T R_{xx}(t - \tau)r_{xx}(\tau, \xi)\, d\tau$$

is transformed into

$$\int_0^T \{Q(p)R_{xx}(t - \tau)\}r_{xx}(\tau, \xi)\, d\tau, \quad (13)$$

and, since $N_{xx}(j\omega)/Q(j\omega)$ is the Fourier transform of $R_{xx}(\tau)$,

$$Q(p)R_{xx}(t - \tau) = N_{xx}(p)\, \delta(t - \tau), \quad (14)$$

which implies that

$$Q(p)\int_0^T R_{xx}(t - \tau)r_{xx}(\tau, \xi)\, d\tau = N_{xx}(p)r_{xx}(t, \xi). \quad (15)$$

Now, since (3) holds for $0 \leq t \leq T$, (15) is valid for $0 < t < T$, and hence

$$Q(p)\int_0^T R_{xx}(t - \tau)r_{xx}(\tau, \xi)\, d\tau$$

and $N_{xx}(p)r_{xx}(t, \xi)$ will differ by, at most, delta functions at the end points. Consequently, operating on both sides of (3) with $Q(p)$ will yield a system of differential equations of the form

$$\mathbf{N}(p)\mathbf{R}^{-1}(t, \xi) = Q(p)\mathbf{I}\, \delta(t - \xi) + \mathbf{\Delta}'(t, \xi), \quad (16)$$

in which $\mathbf{N}(p)$ is a matrix with elements $N_{xx}(p)$, $\cdots$, $N_{yy}(p)$, and $\mathbf{\Delta}'(t, \xi)$ is a matrix with elements

$$\Delta'_{ij}(t, \xi) = \sum_m [D_{ij}^{(m)}\, \delta^{(m)}(t) + E_{ij}^{(m)}\, \delta^{(m)}(t - T)], \quad (17)$$

where the $D_{ij}^{(m)}$ and $E_{ij}^{(m)}$ are undetermined coefficients and $m \leq q - 1$, q being the order of $Q(p)$.

A general solution of (16) may be written as

$$\mathbf{R}^{-1}(t, \xi) = \mathbf{N}^{-1}(p)\{Q(p)\mathbf{I}\, \delta(t - \xi) + \mathbf{\Delta}'(t, \xi)\} + \mathbf{\Lambda}(t, \xi), \quad (18)$$

where $\Lambda(t, \xi)$ is a matrix with elements

$$\Lambda_{ij}(t, \xi) = \sum_\beta A_{ij}^{(\beta)}\varphi_\beta(t), \quad (19)$$

in which the $A_{ij}^{(\beta)}$ are constants and the $\varphi_\beta(t)$ are linearly-independent solutions of the homogeneous system $\mathbf{N}(p)\boldsymbol{\phi} = 0$. Since $\mathbf{\Delta}'(t, \xi)$ contains delta functions of order at most $q - 1$, the term $\mathbf{N}^{-1}(p)\mathbf{\Delta}'(t, \xi)$ will contain delta functions of order at most $q - n - 1$, where n is the degree of the lowest term in $\mathbf{N}(p)$. Consequently, we have

$$\mathbf{R}^{-1}(t, \xi) = Q(p)\mathbf{N}^{-1}(p)\mathbf{I}\, \delta(t - \xi) + \mathbf{\Delta}(t, \xi) + \mathbf{\Lambda}(t, \xi), \quad (20)$$

where $\mathbf{\Delta}(t, \xi)$ is a delta-function matrix

[4] This definition of the probability-density functional $P_T(u, v)$ is based on an analogous interpretation of Preston's result for the case of a single Gaussian process suggested by George Turin. $P_T(u, v)$ can also be defined as a Radon-Nikodym derivative. For related results see E. Parzen, "Statistical inference on time series by Hilbert space methods, 1," Appl. Math. and Stat. Lab., Stanford University, Stanford, Calif., Tech. Rept. 23; January, 1959.

with elements of the form

$$\Delta_{ij}(t, \xi) = \sum_{m=0}^{q-n-1} [B_{ij}^{(m)} \delta^{(m)}(t) + C_{ij}^{(m)} \delta^{(m)}(t - T)], \quad (21)$$

in which the $B_{ij}^{(m)}$ and $C_{ij}^{(m)}$ are undetermined coefficients.

As in the case of the one-dimensional integral equation,[5] the $B_{ij}^{(m)}$ and $C_{ij}^{(m)}$ may be determined by substituting (21) into (3) and treating the resulting equations as identities. To illustrate, consider a simple example in which $\{x(t)\}$ and $\{y(t)\}$ are stationary processes with the spectral-density matrix

$$\mathbf{G}(p) = \frac{1}{a^2 - p^2} \times \cdot \begin{bmatrix} a^2 - p^2 & a - p \\ a + p & a^2 + 1 - p^2 \end{bmatrix} \quad (22)$$

[5] K. S. Miller and L. A. Zadeh, "Solution of an integral equation occurring in the theories of prediction and detection," IRE Trans. on Information Theory, vol. IT-2, pp. 72–75; June, 1956. See also, C. L. Dolph and M. A. Woodbury, "On the relation between Green's functions and covariances of certain stochastic processes and its application to unbiased linear prediction," *Trans. Am. Math. Soc.*, vol. 72, pp. 519–550, May, 1952; and V. S. Pugachev, "Method of solving the basic integral equation of the statistical theory of optimum systems in closed form," *Prik. Mat. i Meh.*, vol. 23, pp. 3–14; January, 1959 (in Russian).

Here det $\mathbf{N}(p) = (a^2 - p^2)^2$ and solutions of the homogeneous equation are $e^{\pm at}$ and $te^{\pm at}$. Thus, a typical term such as $r_{xx}(t, \xi)$ reads

$$r_{xx}(t, \xi) = \delta(t - \xi) + \frac{1}{2a} e^{-a|t-\tau|} + A_{11}^{(1)} e^{-at} + A_{11}^{(2)} te^{-at} + A_{11}^{(3)} e^{at} + A_{11}^{(4)} te^{at} + B_{11}\, \delta(t) + C_{11}\, \delta(t - T). \quad (23)$$

On substituting $r_{xx}(t, \xi), \cdots, r_{yy}(t, \xi)$ into (3) and treating the resulting equation as an identity, one finds the following explicit expressions for the elements of the inverse kernel $\mathbf{R}^{-1}(t, \xi)$:

$$r_{xx}(t, \xi) = \delta(t - \xi) + \frac{1}{2a} e^{-a|t-\xi|} - \frac{1}{2a} e^{-a(t-\xi)} - \frac{1}{a^2 C} \sinh at \sinh a\xi \quad (24)$$

$$r_{xy}(t, \xi) = -e^{-a(t-\xi)} 1(t - \xi) + \frac{1}{aC} \sinh e^{a\xi}\, at \quad (25)$$

$$r_{yx}(t, \xi) = -e^{a(t-\xi)} 1(\xi - t) + \frac{1}{aC} e^{at} \sinh a\xi \quad (26)$$

$$r_{yy}(t, \xi) = \delta(t - \xi) - \frac{1}{C} e^{at} e^{a\xi}, \quad (27)$$

where

$$C = 2ae^{2aT} + \frac{1}{2a} e^{2aT} - \frac{1}{2a}.$$

It should be noted that the determination of $\mathbf{R}^{-1}(t, \xi)$ for higher-dimension vector processes can be carried out in exactly the same manner, but the number of undetermined coefficients in $\mathbf{R}^{-1}(t, \xi)$ increases rapidly with n. One exception is the special case where $\mathbf{N}(p)$ is a constant matrix. For processes of this type, $\mathbf{R}^{-1}(t, \xi)$ is given simply by $\mathbf{R}^{-1}(t, \xi) = Q(p)\delta(t - \xi)$.

J. B. Thomas
Dept. of Elec. Engrg.
Princeton University
Princeton, N. J.

L. A. Zadeh
Dept. of Elec. Engrg.
University of California
Berkeley, Calif.

Contributors

Phillip Bello (S '52—A '55), for bi- raphy, please see page 55 of the January, 51, issue of these TRANSACTIONS.

❖

Jack Capon (S '54—M '56) was born in w York, N. Y., on April 28, 1932. He eived the B.E.E. degree from the llege of the City of New York in 1953, M.S.E.E. degree from the Massachu- ts Institute of Technology, Cambridge, ass., in 1955, and the Ph.D. degree in ctrical engineering from Columbia Uni- rsity, New York, N. Y., in 1959.

He was a research assistant at the search Laboratory of Electronics of .I.T. from 1953 to 1955, and an instructor the Electrical Engineering Department Columbia University from 1955 to 59. At present he is working for the ederal Scientific Corporation, New York, . Y., where he is concerned with the tection and processing of signals in ise, and with problems in spectral alysis.

Dr. Capon is a member of Tau Beta Pi, ta Kappa Nu, Sigma Xi, the Institute of athematical Statistics, the Acoustical ciety of America, and the American ssociation for the Advancement of Science.

❖

Janis Galejs (A '52—M '57) was born Riga, Latvia, on July 21, 1923. He ceived the Engineering Diploma in ectrical engineering from the Technical niversity, Brunswick, Germany, in 1950, d the M.S. and Ph.D. degrees in electrical gineering from the Illinois Institute of echnclogy, Chicago, in 1953 and 1957, spectively.

While attending I.I.T., he worked for e Cook Research Laboratory on fire ntrol problems, radar, and communica- on systems. In 1957 he joined the Applied esearch Laboratory of Sylvania Electric roducts, Inc., Waltham, Mass., where he engaged in studies of radar systems and statistical analysis of radar and com- unication problems. More recently, he s been concerned with propagation in spersive media and in ionosphere.

Dr. Galejs is a member of Sigma Xi and au Beta Pi.

E. M. Glaser (S '49—A '50—M '55) was born in New York, N. Y. on October 17, 1927. He received the B.E.E. degree from The Cooper Union, New York, in 1949. He received the M.S.E. degree in 1954, and the D. Eng. degree in 1960, both from The Johns Hopkins University, Baltimore, Md.

From 1950 to 1952, he was associated with the Glenn L. Martin Company of Baltimore as an electromechanical engineer. From 1952 to 1960, he was a member of the staff of the Radiation Laboratory of The Johns Hopkins University. There he worked primarily in the field of communication theory. Since June of 1960 he has been a Fellow in the Department of Physiology of The Johns Hopkins School of Medicine, where he is engaged in the application of communication theory to problems in the field of neurophysiology.

❖

Marcel J, Golay (SM '51—F '60) was born in Neuchatel, Switzerland, on May 3, 1902. He attended the Gymnase Scientifique of Neuchatel where he received the B.S. degree in 1920, and the Federal Institute of Technology in Zurich, where he received the Licentiate in Electrical Engineering in 1924. He attended the University of Chicago, Chicago, Ill., where he obtained the Ph.D. degree in physics in 1931.

From 1924 until 1928, he was at the Bell Telephone Laboratories. After a short association with the Automatic Electric Company, Chicago, Ill., he entered the civil service in 1931, and was a member of the Signal Corps Engineering Laboratories at Fort Monmouth, N. J., until 1955. He is now serving as consultant to the Philco Corporation, Philadelphia, Pa., and to the Perkin-Elmer Corporation, Norwalk, Conn.

Dr. Golay is a member of the American Physical Society, the Optical Society of America, the American Rocket Society, and the Society for Applied Spectroscopy. He is the recipient of the 1951 IRE Harry Diamond Award and of the 1961 ACS Sargeant Award.

❖

William F. Higgins (S '49—A '50 — M '55—SM '60) was born on September 2, 1920 in Boston, Mass. He received the B.S. degree in 1950, and the M.S. degree in 1953, both in electrical engineering from the University of Massachusetts, Amherst, Mass. During World War II, he served as a communication officer in the Air Force.

From 1950–52 he was an engineer with the Vitro Corporation of America, Silver Spring, Md. He was instructor in electrical engineering from 1951 to 1953 at the University of Massachusetts, and from 1953 to 1955 at Brown University, Providence, R. I., where he was a part-time graduate student. In 1955 he joined the Applied Research Laboratory of Sylvania, Waltham, Mass., where he was an engineering specialist. In 1960 he joined Aeronutronic, a Division of the Ford Motor Company, Newport Beach, Calif., and is presently working on the sensor and communication problems of space surveillance. In 1961 he became a staff member of Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass.

Mr. Higgins is a member of Sigma Xi and Tau Beta Pi.

❖

David Middleton (S '42—A '44—SM '53—F '59), for biography, please see page 413 of the June, 1960, issue of these TRANSACTIONS.

❖

Joseph S. Wholey was born in Cranston, R. I., on February 6, 1935. He received the B.A. degree in mathematics from Catholic University, Washington, D. C., in 1956, and the M.A. degree in mathematics from Harvard University, Cambridge, Mass., in 1958. He is a candidate for the Ph.D degree in philosophy at Harvard University and is currently completing his thesis in mathematic logic on the decidability of axiomatic theories.

In 1955 and 1956 he was employed in the analyzing missile test data at the Applied Physics Laboratory of Johns Hopkins University, Silver Spring, Md. Since 1957 he has combined work at the Applied Science Division of Melpar, Inc., Watertown, Mass., with teaching at the Newton College of the Sacred Heart, Newton, Mass., where he is an assistant professor of mathematics. His work at Melpar includes the application of information theory and approximation methods to pictorial data compression by computers. He has also studied statistical methods of solving electromagnetic reconnaissance problems.

Mr. Wholey is a member of Phi Beta Kappa and Sigma Xi.

Abstracts

This Section of the issue is devoted to abstracts of material which may be of interest to PGIT members. Sources are Government, Industrial and University reports, and books and journals published outside of the United States. Readers familiar with material of this nature which is suitable for abstracting are requested to communicate the pertinent information to one of the Editors or Correspondents listed below.

Zeros of a Random Stationary Signal—H. Debart (in French). (*Cables & Transmission*, vol. 14, pp. 191–199; July, 1960.)

Consider a random stationary signal $Y(t)$; then if $x(t)$ is set equal to 1 for $Y(t) > 0$ and equal to -1 for $Y(t) < 0$, the study of the zeros of $Y(t)$ is equivalent to the study of the transitions of $x(t)$. Several known results are stated, and subsequently some characteristics of the distribution of the zeros of $x(t)$ are obtained by expanding $x(t)$ in the Loéve-Karhunen expansion. The results are approximate, but relatively easy for numerical computations.

A Systematic Code for Non-Independent Errors—T. Kasami (in Japanese). *J. Information Processing Soc. Japan*, vol. 1, pp. 132–137; November 3, 1960.)

A class of systematic codes is described; these codes are designed to correct any one of the following code errors: single error, double-adjacent error, three-binit-wide double error, and triple-adjacent error. It is shown that the codes considered are highly efficient. A pair of linear feedback shift registers may be used for the purpose of constructing this class of codes.

Let the "code-word length" and the "number of check digits" be denoted by n and m, respectively. Then for an even number m, the "complete" codes are given whose parity-check matrices are constructed by using two sequences of the following type: a maximum-length sequence of period 3 and a suitable maximum-length sequence of period $2^{m-2} - 1$. These codes are equivalent to those obtained by Melas. The condition is then considered by which the maximum-length sequence of period $2^{m-2} - 1$ should be selected and, as an example, a simple decoding procedure is also presented.

For an odd number m, a method is proposed which permits the systematic construction of codes. For example, this method yields a (27, 20) code and a (121, 112) code, both of which are more efficient than the respective Reiger code and are as easily realized by electronic devices.

Phonetic Recognition and its Measurement—J. C. Lafon (in French.) (*Ann. des Télécommunications*, vol. 15, pp. 27–37; January–February, 1960.)

A "phoneme" is tentatively defined as being the smallest phonetic unit which permits the distinction of two words of different meaning, differentiated just by that acoustic unit. Phonetic integration represents the understanding by sensory means of acoustic symbols called phonemes and the possibility of distinguishing the criteria necessary for their individualization. First, understanding is studied—that is, the development of the phonetic complex and its acquisition by a child; subsequently, the methods of measurement are studied. Different perturbations are evoked in terms of their peripheral or central localization. Finally, the neurological and phonetic aspects are considered and the practical applications of their measurement.

On a New Theory of the Limitation of Signal Spectra—J. Oswald (in French). (*Cables & Transmission*, vol. 14, pp. 249–261; October, 1960.)

The object of this study is a critical examination of the present theory of the limitation of signal spectra. The well-known difficulties which arise from the strict limitation of spectra, in particular the simultaneous localization of a signal in time and in a frequency interval, come from an erroneous interpretation concerning the operation of spectral limitation. The solution presented here circumscribes all these difficulties; it permits the establishment of a general and coherent theory of the limitation of spectrum and, subsequently, of determining the signal transformation by ordinary operators (especially those that are associated with linear networks) without specifying them. It is therefore possible to consider ideal operations of filtering, integration, differentiation, etc., the results of which are in perfect agreement with those of the rigorous theory applied to particular instances (composition products), or even with the elementary conclusions of common sense.

The essential idea which is the basis of this work is the substitution of a distribution for the continuous functions generally utilized to represent signals and operators; it can be concluded that, at the cost of an amplitude quantization, all continuous linear operators can be replaced by arithmetic or digitized operators.

A Method of Finding the Original Message as Accurately as Desired From a Finite Number of Observations After a Rectangular Band-pass Filter—H. Wolter (in German). (*Arch. Elekt. Übertragung*, vol. 13, pp. 393–404; 1959.)

If a finite message is observed with a device providing an extremely sharp cutoff of the frequency band, and a calculation of the original message from it is demanded with an error $< \epsilon$, a measuring error bound $\delta(\epsilon) > 0$ always exists in such a way that the original message can be calculated with the required accuracy from a finite number of observations with errors $< \delta$. The proof uses a method of summation of divergent series due to Euler. It is essential that one know the duration of the original message.

On the Limiting Behavior of Extremely Selective Communication Channels in Information Theory—H. Wolter (in German). *Arch. Elekt. Übertragung.*, vol. 13, pp. 171–174; 1959.)

A Gaussian error function cannot be the frequency function of an information channel. A sequence of filters can approximate the Gaussian behavior in the amplitude channel, but then the phase

aracteristic diverges. Therefore the use of Gaussian pulseforms d characteristics in information rate calculations is inadmissible.

e Fundamental Theorems of Information Theory as a Conse- ence of Error Propagation in the Solution of Convolution Inte- als—H. Wolter (in German). (*Arch. Elekt. Übertragung.*, vol. , pp. 101–113; 1959.)

Given a filter (channel), a random telegraph signal, and the dition of white noise, one asks for the optimum speed at which e signal should be sent through the channel. If the length of a digit τ seconds, the capacity is calculated as $C = 1/\tau = (\alpha/2\pi)(N_s/n_r)$ is a constant of about 1, N_s is the signal power, and n_r is the ise power per unit bandwidth). The criterion chosen is that at e optimum speed the mean square error due to noise and that e to distortion are both half the mean square of the signal. (Thus pacity is not used herein in the sense of the coding theorem.) ne next question is: if the signal is quantized in m steps (instead 2), what is then the best speed? Since the accuracy required es up by m^2, the capacity goes down by $(2 \ln m)/(m^2/4)(2 \ln m$ is e gain per time unit due to m steps).

n the Fundamental Theorem of Information Theory, Particularly pplied to Optics—H. Wolter (in German). (*Physica*, no. 24, pp. 7–475; 1958.)

In optics there exists a theorem analogous to the Nyquist theorem. pecifically, it is impossible to know the angle α_x from which photons rive with an accuracy better than $\Delta(\sin \alpha_x) = \lambda/\Delta x$ (λ is the avelength and Δx the aperture width). However, the author shows at by special means, *e.g.*, the use of crossed Fresnel biprisms, gain of about 300 in angular accuracy can be reached. The limi- tion is then given by the number N of photons available $\times . \Delta\alpha_x/\lambda = \lambda/\sqrt{N}$.

For the measurement of optical grids with insufficient aperature, e introduction of a second grid nearly parallel to the first gives ough information in the image to deduce the otherwise unavail- le constants.

n the Fundamental Theorems of Information Theory, Particularly pplied to Communications—H. Wolter (in German). (*Arch. Elekt. Übertragung.*, vol. 12, pp. 335–345; 1958.)

According to the Nyquist (Kupfmuller) criterion, resolution time and bandwidth are related by $\Delta t \Delta v \leq \frac{1}{2}$. Shannon's sampling eorem states that any function limited to the bandwidth W and me interval T can be specified by giving $2TW$ numbers. However, mmunication with exactly limited bandwidth is impossible (the aley-Wiener criterion; a less exact proof is given herein). From tical and communication examples, it is shown that where the andwidth is not limited so exactly, the resolution can be made uch better than expected from the above relations by equalization ompensation filters). The only limitation is then effectuated by ise.

The following papers appear in the Transactions of the First Prague onference on Information Theory, Statistical Decision Functions, nd Random Processes (held on November 28–30, 1956). These Trans- ctions were published by the Publishing House of the Czechoslovak cademy of Sciences, Prague, Czechoslovakia, 1957. The affiliations the authors are given below; abstracts are given when available.

he Entropy of Functions of Finite-State Markov Chains—D. lackwell (in English). (University of California, Berkeley, Calif.)

It is shown that the entropy H of an ergodic process $y_n, -\infty < n < \infty\}$ with states $a = 1, \cdots, A$ such that $y_n = \Phi(x_n)$ here $\{x_n\}$ is a stationary ergodic finite-state Markov process with ates $i = 1, \cdots, I$ and transition matrix $M = || m(i, j) ||$ is given by

$$H = -\int \sum_a r_a(w) \log r_a(w) \, dQ(w),$$

here r_a is a function, defined on the set W of all $w = (w_i, \cdots, w_I)$ ch that

$$w_i \geq 0, \quad \sum_i^I w_i = 1, \quad r_a(w) = \sum_{i,j \ni \Phi(j) = a} w_i m(i, j),$$

nd Q is the distribution of the conditional distribution of x_0 given $y_0, y_{-1}, \cdots$. An integral equation is obtained for Q, and a method is given for showing, under rather strong hypotheses, that the solution of this integral equation is unique. An example in which Q is singular is given.

On Some Soviet Work in Information Theory—B. V. Gnedenko (in Russian). (Math. Inst., Ukrainian Acad. Sci., Kiev, USSR.)

A Display of Information Theory Problems Concerning Telephone Transmission—H. Hansson (in English). (Tel. AB L. M. Ericsson, Sweden.)

The Selectivity of Parametric Tests—C. Rajski (in English). (Inst. Math., Polish Acad. Sci., Warsaw, Poland.)

Let Q be the unknown value of a parameter of a general population whose distribution function is known; let n be the sample size and Ω— the critical region; The possible results of testing the hypothesis stating that the actual value of the parameter is Q is described, in a statistical sense, by the power function of the test denoted here by $M(\Omega, n, G)$. The power function may be "better" or "worse," the judgement being usually based on two values taken by the power function for the values of Q assumed in the null hypothesis (Q_0) and in the alternative hypothesis (Q_1). As usual, we write

$$M(\Omega, n, Q_0) = x,$$

$$M(\Omega, n, Q_1) = 1 - \beta.$$

Any monotonically increasing function of x and β, say $r(x, \beta)$, can serve as a measure of "goodness" of the power function of the test, the smaller values of $r(x, \beta)$ indicating that the test is a "better" one.

This method is rather unsatisfactory, as the evaluation is based on two points only of the power function. A more elaborate qualification can be obtained by treating the unknown parameter as a random variable. Its entropy denoted here by $L(\Omega, n)$ and, defined by the formula

$$L(\Omega, n) = -\int \frac{\partial M}{\partial Q} \log \frac{\partial M}{\partial Q} \, dQ,$$

seems to be a more suitable measure of "goodness" of the test as based on the shape of the power function as a whole.

The Bayes Rule and Entropy—C. Rajski (in English). (See above for affiliation.)

By the Bayes rule we mean the assumption that in the lack of the empirical knowledge concerning the *a priori* distribution of the unknown parameter lying in the finite range, this distribution should be taken as a uniform one. The severe criticism of this assumption is widely known. Here the attempt will be made to support the Bayes rule and to present the extension of this rule to the cases of semi-infinite and infinite ranges.

Remarks on Linear Prediction by Means of a Learning Filter—L. Prouza, (in German). (Res. Inst., for Radio Engrg., Pardubice, Czechoslovakia.)

Continuous Random Decision Processes Controlled by Experience—M. Driml and A. Špaček (in English). (Inst. of Radio Engrg. and Electronics, Czechoslovak Acad. Sci., Prague, Czechoslovakia.)

This paper contains a generalization of the theory of experience to the "continuous" parameter case. Roughly speaking, there is given a generalized random process depending on the continuous time parameter and the "value" of which at each time instant is a statistical decision problem. Under proper assumptions it is possible to choose a decision process depending on the time parameter as well, such that the average risk defined conveniently converges to the least possible constant or to a limit which lies in a given neighborhood of this minimum. The construction of this time-dependent decision process is controlled automatically by the experience obtained by storing in a proper way the past values of the process.

Generalized Random Variables—O. Hanš (in English). (Inst. of Radio Engrg. and Electronics, Czechoslovak Acad. Sci., Prague, Czechoslovakia.)

Rankom Fixed-Point Theorems—O. Hanš (in English). (See above for affiliation.)

Inverse and Adjoint Transforms of Linear Bounded Random Transforms—O. Hanš (in English). (See above for affiliation.)

The inverse and adjoint transforms of linear transforms mapping some part of a Banach space into another Banach space are useful tools for studying various problems of functional analysis. It seems reasonable to deal with similar questions for linear random transforms. In this paper some measurability problems for inverse and adjoint transforms of linear bounded random transforms are solved.

Almost-Sure Convergence Theorem for Random Schwartz Distributions—O. Hanš (in English). (See above for affiliation).

Note on Generalized Random Variables—J. Nedoma (in English). (Inst. of Radio Engrg. and Electronics, Czechoslovak Acad. Sci., Prague, Czechoslovakia.)

The Capacity of a Discrete Channel—J. Nedoma (in English). (See above for affiliation.)

Research on the problem of the transmission of discrete messages has so far proceeded in two principal directions. On the one hand, it has been concerned with the question of the extent to which a text with certain statistical properties could be abbreviated by coding. The other problem considered by some authors is the question of minimized noise in the transmission of messages. For this purpose it was necessary to introduce a suitable mathematical model and also certain quantitative characteristics describing the communication link in terms of those properties that are essential in the transmission of messages.

The problem of a suitable mathematical model, *e.g.*, for a transmission channel, has apparently not yet been finally solved, even in the case of a discrete message. This is borne out by the fact that in four significant papers on this subject, namely the papers by Shannon, McMillan, Feinstein and Khintchine, the concept of a channel is defined in different ways. The kernel of these papers is the question of the validity of the Fundamental Shannon Theorem, which McMillan formulates in the following way.

Let the given channel have capacity C and the given source have rate H. If $H < C$, then given any $\epsilon > 0$, there exists an integer $n(\epsilon)$ and a transducer (depending on ϵ) such that when $n(\epsilon)$ consecutive received letters are known, the corresponding n transmitted letters can be identified correctly with probability at least $1 - \epsilon$. If $H > C$ no such transducer exists. McMillan draws attention to the fact that the proof of this theorem requires the channel to be in some sense "continuous."

The present paper does not define the concept of a channel as broadly as is done in McMillan's paper, on which the present paper is mainly based. Nevertheless, though our restriction entails a definite continuity, it turns out that even in this case the above-mentioned Shannon Theorem may not be valid. The results of this paper will be discussed in Chapter V.

The paper is subdivided into five chapters, the first two of which are mainly concerned with the formulation of the required concepts and proofs of some of their properties. The main subject of Chapter III is the proof of Shannon's Theorem for channel capacity defined with respect to the probability of error or to the average frequency of error respectively and also the proof of the equality of both these capacities. Chapter IV investigates the relation between the probability-of-error capacity and Shannon's channel capacity.

The notation used in the paper follows for the most part McMillan's notation. The majority of basic theorems have an asymptotic character and the conditions imposed on the basic concepts enable us to apply the analysis of the infinite-dimensional case also to the finite-dimensional case (for "sufficiently large n"). Therefore, this paper introduces for a number of concepts both "infinite-dimensional" and "finite-dimensional" versions (*e.g.*, sequences and n-dimensional vectors, integral and summation forms of certain characteristics, etc.). The corresponding version is then used as required.

Generalized Concepts of Uncertainty, Entropy and Information from the Point of View of the Theory of Martingales—A. Pérez (in French). (Inst. of Radio Engrg. and Electronics, Czechoslovak Acad. Sci., Prague, Czechoslovakia.)

On the Theory of Information in the Case of an Abstract Alphabet—A. Pérez (in French). (See above for affiliation.)

On the Convergence of Uncertainties, Entropies and Sampled Information to their True Values—A. Pérez (in French). (See above for affiliation.)

An Elementary Experience Problem—A. Špaček (in French). (See above for affiliation.)

Extensions of Random Transformations—A. Špaček (in French). (See above for affiliation.)

Some Theorems on Random Schwartz Distributions—M. Ullrich (in English). (Inst. of Radio Engrg. and Electronics, Czechoslovak Acad. Sci., Prague, Czechosolovakia.)

A Theorem on Extremes of Entropy—L. Votavová (in German). (Inst. of Radio Engrg. and Electronics, Czechoslovak Acad. Sci., Prague, Czechoslovakia.)

Suppose that the largest probability value of a discrete random variable is fixed, but that some of the probabilities are unknown. The proof is given that maximum entropy occurs when the lacking probabilities are assumed to be equal. Minimum entropy is found by setting as many as possible of the unknown probabilities equal to the largest known, by setting one probability equal to the complement with unity of the sum of all former probabilities, and equating, in consequence, the remaining probabilities to zero.

Experience in Games of Strategy and in Statistical Decisions—K. Winkelbauer (in English). (Inst. of Radio Engrg. and Electronics, Czechoslovak Acad. Sci., Prague, Czechoslovakia.)

This paper is devoted to a group of problems called problems of the theory of experience. The concept of experience is described, and the motives which have led to the choice of the mathematical model for statistical decision used herein are stated. It is shown that the mathematical model of a communication system is identical with that of statistical decision problems. The performance of the system under assumed knowledge of the *a priori* probability distribution of transmitted signs is defined and called the average risk. Its lower bound is called the Bayes risk. The statistical decision functions may be chosen in such a way that the average risk converges toward the Bayes risk. The same, however, is true even in the case when the *a priori* distribution is unknown at the receiver, provided that it is constant. One has to replace only the unknown probability distributions obtained in a prescribed manner from the experience gained in preceding steps.

In part, a somewhat different methodical approach to the basic concepts of the theory of games is taken and is developed in an expository manner.

ook Reviews

sting Statistical Hypotheses—E. L. Lehmann. (John Wiley and ns, Inc., New York, N. Y.; 1959. 369 Pages. $11.00)

There are several areas in which the techniques of mathematical tistics can be useful to a radio or electronics engineer, *e.g.*, quality ntrol in component production, design of experiments and inter-etation of results in propagation-research, and design of radio radar receivers to work effectively under adverse conditions of ise, clutter, multipath, etc. Presumably, however, it is the last-entioned application which has really stimulated an interest in e theory of statistical inference on the part of radio engineers d which, indeed, prompts a review of a book of this kind in this urnal. It seems fair, then, to observe at the outset that, because its coverage, Professor Lehmann's book has limited usefulness the solution of current problems in statistical decision theory ising in radio communications, radar, radio astronomy, and ied fields. The book does not cover parameter estimation nor e kind of theory which arises in the application of statistical ference to stochastic processes; both of which are necessary in e class of problems referred to. The book does provide an excellent eatment of one part of statistical decision theory (the small-mple theory of hypothesis testing) which would certainly be lpful to a person who is developing a thorough grounding on hich to base his applied work.

Chapter 1, entitled "The General Decision Problem," contains preliminary discussion of the formulation of a decision problem terms of sample space, parameter space and decision space. The ncepts of loss and risk and optimum procedures are introduced, cluding Bayes and minimax procedures. The maximum-likelihood ethod is discussed briefly, and the chapter closes with an informa-ve, non-measure-theoretic introduction to sufficient statistics. he reviewer feels that this chapter is the best introduction to odern statistics he has read and recommends it particularly to e nonstatistician who wants to get some general feeling for what odern statistics is about. The second chapter is a rather brief view of the appropriate parts of measure and probability theory, d will probably provide rough going for one unfamiliar with easure-theoretic probability.

With Chapter 3 the book gets to its central theme, as stated the title. The problem of testing a simple hypothesis against a mple alternative is introduced and the fundamental lemma of Neyman and Pearson, leading to the likelihood-ratio test, is carefully stated and proved. An immediate extension is made to the particular class of problems, in which a compound hypothesis is tested against a compound alternative, for which the family of probability densities possesses the property of monotone likelihood ratio. In this case, likelihood ratio tests are UMP (uniformly most powerful). Some additional topics in Chapter 3 are: an extension of the Neyman-Pearson lemma to more side conditions, a discussion of confidence bounds, and a proof of the optimality of the sequential probability ratio test.

The central difficulty in the theory of hypothesis testing and the thing that makes the subject nontrivial is, of course, the fact that in general, with compound hypotheses and alternatives, there exists no UMP test. The situation the statistician has faced has been to devise consistent, convincing test procedures to work in problems for which by the simple power criterion no best test exists and, indeed, where by the same criterion two tests often can not even be compared. What has been done has been to impose reasonable restrictions on the class of tests allowed and then to search for UMP tests within these restricted classes. The usual restrictions are that the test be *unbiased,* satisfy a *similarily* condition, or be *invariant* under some group of transformations. In this book a thorough treatment of unbiased and similar tests is given in Chapters 4 and 5 and of invariant tests in Chapter 6. Chapter 7 is a long chapter on linear hypothesis testing (of which linear regression is a special case). Chapter 8 develops some theory using the minimax principle, which provides another and potentially quite general way of getting around the difficulty of no UMP test.

The book has many examples worked out in the text, and an extensive list of problems for the reader. The examples include most of the standard testing problems involving Gaussian distributions and the well-known derived distributions, such as the ψ^2-distribution or Student's t-distribution; and a worker in communications theory might find these immediately useful. There are also nonparametric examples. Each chapter has a fairly lengthy bibliography.

William L. Root
M.I.T. Lincoln Lab.
Lexington, Mass.

A STATEMENT OF EDITORIAL POLICY

The IRE Transactions on Information Theory is a quarterly journal devoted to the publication of papers on the transmission, processing, and utilization of information. The exact subject matter of acceptable papers is intentionally, by editorial policy, not sharply delimited. Rather, it is hoped that as the focus of research activity changes, a flexible policy will permit the Transactions to follow suit and that it will continue to serve its readers with timely articles on the fundamental nature of the communication process. Topics of current appropriateness include the coding and decoding of digital and analog communication transmissions, studies of random interferences and of information bearing signals, analyses and design of communication and detection systems, pattern recognition, learning, automata, and other forms of information processing systems.

Papers can be of two kinds, tutorial or research, and should be so indicated. The former must be well-written expositions summarizing the state of a field in which research is still in progress, or else unifying results scattered in the literature. Research papers must be original contributions not published elsewhere. They must present new methods, concepts, or ideas, or extend old ones to new areas of applicability; or, they must present new data, findings or inventions, or solve new problems of more than casual interest. They will not be accepted if, in the view of the reviewers and editors, they constitute a straightforward and easy application of existing theory to a special case of limited interest. It is not necessary that the length of each research paper be great; on the contrary, the submission of short but formal research notes is to be encouraged.

In addition to papers, readers are invited to submit notes to the Correspondence section. These may include such things as early summaries of important work to be published later at greater length, or remarks on material that has already appeared. Contributions in the form of "problem statements" are also sought for the Correspondence section. This category includes problems to which the author knows no solution but suspects that another reader might, conjectures for which a proof or disproof is desired, and so forth.

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Editor for possible publication in the PGIT Transactions. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or submitted for publication elsewhere.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in India ink or drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Papers should be prepared for publication in a matter similar to those intended for the Proceedings of the IRE. Further instructions may be obtained from the Editor. The original copy and drawings of material not accepted for publication will be returned.

All technical manuscripts and editorial correspondence should be addressed to Arthur Kohlenberg, Melpar, Inc., 11 Galen Street, Watertown 72, Mass.

Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to the PGIT Newsletter, c/o Prof. N. M. Abramson, Electrical Engineering Department, Stanford University, Stanford, Calif.

INSTITUTIONAL LISTINGS